Eco-evolutionary dynamics for finite populations and the noise-induced reversal of selection

Ananda Shikhara Bhat[1,2,3,*], Vishwesha Guttal[1]

[1] Centre for Ecological Sciences, Indian Institute of Science, Bengaluru, Karnataka-560012, India
[2] Department of Biology, Indian Institute of Science Education and Research Pune, Maharashtra-411008, India
[3] *Current Affiliation:* Institute of Organismic and Molecular Evolution (iomE), Johannes Gutenberg University, 55128 Mainz, Germany; Institute for Quantitative and Computational Biosciences (IQCB), Johannes Gutenberg University, 55128 Mainz, Germany

∗ Corresponding author. E-mail: abhat@uni-mainz.de

Short title: Evolution in finite populations

Keywords: Eco-evolutionary dynamics, finite populations, demographic stochasticity, noise-induced selection, bet-hedging, population biology

Abstract

Theoretical studies from diverse areas of population biology have shown that demographic stochasticity can substantially impact evolutionary dynamics in finite populations, including scenarios where traits that are disfavored by natural selection can nevertheless increase in frequency through the course of evolution. Historically, most general analytic frameworks have either restricted themselves to models with constant or deterministically varying total population size or have resorted to dynamically insufficient formulations. Here, we analytically describe the eco-evolutionary dynamics of finite populations from demographic first principles to investigate how noise-induced effects can alter the evolutionary fate of populations in which total population size may vary stochastically over time. Starting from a generic birth-death process describing a finite population of individuals with discrete traits, we derive a set of stochastic differential equations (SDEs) that recover well-known descriptions of evolutionary dynamics such as the replicator-mutator equation, the Price equation, and Fisher's fundamental theorem in the infinite population limit. For finite populations, our SDEs reveal how stochasticity can induce a directional evolutionary force termed 'noise-induced selection' via two distinct mechanisms, one that operates over relatively faster (ecological) timescales and another that is only apparent over longer (evolutionary) timescales. Despite arising from the stochasticity of finite systems, the effects of noise-induced selection are predictable and may oppose natural selection. In some cases, noise-induced selection can even reverse the direction of evolution predicted by natural selection. By extending and generalizing some standard equations of population genetics, we thus describe how noise-induced selection appears alongside and interacts with the more well-understood forces of natural selection, neutral drift, and transmission effects (mutation/migration) to determine the eco-evolutionary dynamics of finite populations of non-constant size.

Introduction

Eco-evolutionary population biology has a strong mathematical underpinning, and can broadly be captured mathematically under a single unifying framework via the replicator-mutator equation and the Price equation. This formalism yields several relevant mathematical structures of evolution, such as evolutionary game theory and classic population genetics, as special cases (Page and Nowak, 2002; Queller, 2017; Lion, 2018). The Price equation partitions changes in population composition into multiple terms, each of which lends itself to a straightforward interpretation in terms of the high-level evolutionary forces of selection and mutation, thus providing a useful mathematical framework for describing how populations change over time (Frank, 2012). The Price equation also leads to a number of simple yet insightful 'fundamental theorems' of population biology and unifies several various seemingly disjoint formal structures under a single theoretical banner (Queller, 2017; Lion, 2018; Lehtonen, 2020; Luque and Baravalle, 2021). However, the replicator-mutator equation, Price equation, and related 'fundamental theorems' of evolutionary dynamics are usually formulated in a deterministic setting that neglects stochastic fluctuations due to finite population effects (Page and Nowak, 2002; Queller, 2017; Lion, 2018).

Today, we increasingly recognize that incorporating the finite and stochastic nature of the real world routinely has much stronger consequences than simply 'adding noise' to deterministic expectations and can cause qualitative changes in the behavior of diverse biological systems (Horsthemke and Lefever, 1984; Black and McKane, 2012; Boettiger, 2018; Jhawar et al., 2020; Majumder et al., 2021; DeLong and Cressler, 2023; Yamamichi et al., 2023). In ecology and evolution, stochastic models need not exhibit phenomena predicted by their deterministic analogues (Proulx and Day, 2005; Johansson and Ripa, 2006; Black and McKane, 2012; Débarre and Otto, 2016). They may also exhibit novel phenomena not predicted by deterministic models (Constable et al., 2016; Rogers and McKane, 2015; Joshi and Guttal, 2018; DeLong and Cressler, 2023).

A striking example of such novel phenomena is the complete 'reversal' of the evolutionary trajectory that is seen in some finite population evolutionary models (Houchmandzadeh and Vallade, 2012; Constable et al., 2016; McLeod and Day, 2019a; Mazzolini and Grilli, 2023). For example, in public goods games, the production of a costly public good is susceptible to invasion by 'cheaters' who use the public good but do not produce it. Due to this, standard (deterministic) evolutionary game theory predicts that producers should eventually go extinct. However, in finite, fluctuating populations, producers not only persist but also outcompete non-producers, the exact opposite of infinite population predictions (Constable et al., 2016; McLeod and Day, 2019a). This phenomenon of evolution proceeding in the direction of the classically disfavored type that leads to the 'reversal' of the prediction of deterministic natural selection has been dubbed 'noise-induced selection' (Week et al., 2021). Such noise-induced effects have been seen in several models in fields as diverse as epidemiology (Humplik et al., 2014; Parsons et al., 2018; McLeod and Day, 2019b; Day et al., 2020), cell-cycle dynamics (Wodarz et al., 2017), and social evolution (Houchmandzadeh and Vallade, 2012; Chotibut and Nelson, 2015; Constable et al., 2016; McLeod and Day, 2019a).

Despite the ubiquity of the phenomenon of qualitative noise-induced effects on evolutionary trajectories, we currently lack a description of how classic equations of evolution such as the replicator-mutator equation and Price equation are affected by such demographic stochasticity. Bet-hedging theory, a branch of evolutionary ecology that aims to build general theories that capture the effects of stochasticity on eco-evolutionary dynamics (Seger and Brockmann, 1987; Frank and Slatkin, 1990; Starrfelt and Kokko, 2012), has typically worked with abstract populations in which the mean reproductive output and the variance in reproductive output are essentially independent parameters (Gillespie, 1974; Gillespie, 1977; Frank and Slatkin, 1990; Shpak, 2005). Furthermore, models of bet-hedging have typically worked with populations affected by spatiotemporally fluctuating *external* noise that is the result of stochastic fluctuations in external environments (Seger and Brockmann, 1987; Olofsson et al., 2009; Childs et al., 2010; Starrfelt and Kokko, 2012), whereas most formulations of the standard equations of population genetics (Page and Nowak, 2002; Lion, 2018; Lehtonen, 2018) as well as many models showing noise-induced effects on eco-evolutionary dynamics (Parsons et al., 2010; Houchmandzadeh and Vallade, 2012; Constable et al., 2016; Parsons et al., 2018; McLeod and Day, 2019a; Day et al., 2020) do not model any external environment at all. Due to this, it is often unclear *a priori* under what situations these noise-induced effects become important for evolutionary dynamics or how these effects interact with the more well-understood evolutionary forces of natural selection, mutation, and drift (Yamamichi et al., 2023). For example, how does noise-induced selection interact with genetic drift, or indeed natural selection? Are 'noise-induced selection' and 'bet-hedging' essentially the same effect that has been spoken about using different terminology, or are there multiple distinct phenomena at play?

Stochastic individual-based models are a natural choice for describing the dynamics of finite populations to try and examine the interplay of noise-induced effects and more well-understood evolutionary forces using a first principles approach. Here, probabilistic rules for birth and death are specified at the individual level. Such models allow us to capture a stochastically varying population size, and thus enable us to relax assumptions of constant population size as seen in models such as the Wright-Fisher or Moran process (Lambert, 2010; Abu Awad and Coron, 2018). From these individual-based rules, we can now systematically derive population-level dynamics and thus avoid potential pitfalls that can arise when simply adding noise terms to a 'deterministic skeleton' in an ad-hoc fashion (Coulson et al., 2004; Black and McKane, 2012; Strang et al., 2019). Furthermore, since demographic processes such as birth and death rates explicitly account for the ecology of the system, they can more accurately reflect the complex interplay between ecological and evolutionary processes and provide a fundamental, mechanistic description of the relevant population dynamics (Lambert, 2010; Doebeli et al., 2017).

In this paper, we derive general equations for the dynamics of finite, fluctuating populations evolving in continuous time starting from such mechanistic first principles (Fig 1). These equations reduce to well-known results such as the replicator-mutator equation and the Price equation in the infinite population limit, thus illustrating consistency with the known formal structures of eco-evolutionary population dynamics (Queller, 2017; Lion, 2018). For finite populations, these same equations also provide a generic description and synthesis of the noise-induced effects of finite population size and their consequences for eco-evolutionary population dynamics. Such a systematic derivation provides relations between ecological quantities such as the expected population growth rate and the variance in population growth rate and connects them to evolutionary forces such as natural selection and genetic drift in trait frequency space. Using these equations, we synthesize the connections between noise-induced effects on population dynamics, including the 'Gillespie effect' of bet-hedging theory (Gillespie, 1977), 'noise-induced effects' in ecological population models (Constable et al., 2016; Parsons et al., 2018), 'drift-induced selection' (Veller et al., 2017; Saunders et al., 2018), 'noise-induced selection' (Week et al., 2021), and long-term effects of demographic stochasticity through the effects of 'evolutionary noise' (McLeod and Day, 2019a; McLeod and Day, 2019b).

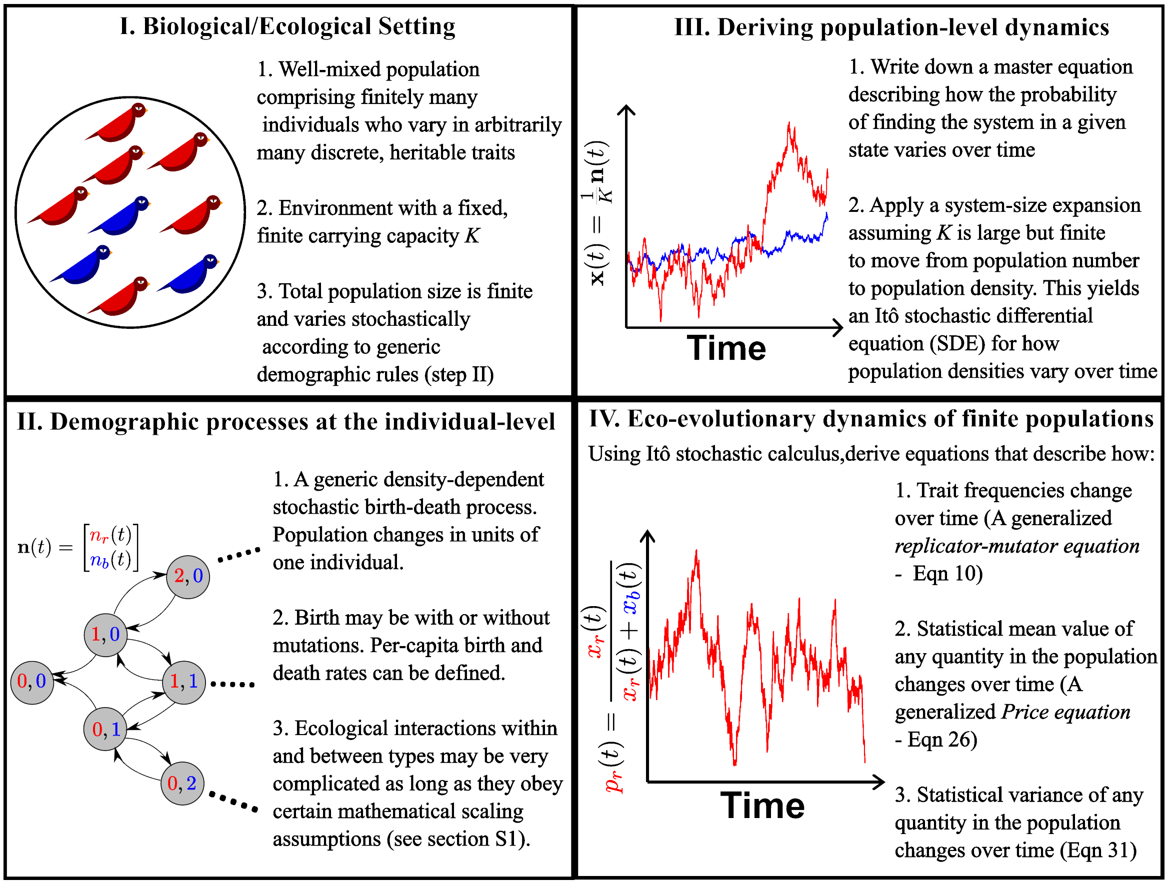

Figure 1: An outline of the approach we adopt in this paper

A stochastic birth-death process for population dynamics

We consider a well-mixed population that can contain up to m different types of individual entities. For example, a gene may have m different alleles, individuals within a species may come in one of m phenotypes, or a community may have m different species; we refer to each distinct variant of an entity as a 'type'. Unlike many classic stochastic formulations in evolutionary theory (Crow and Kimura, 1970; Lande, 1976; Kimura and Ohta, 1974), we do not assume a fixed or deterministically varying (effective) population size. Instead, we allow the total population size to emerge naturally, and thus fluctuate stochastically, from the stochastic birth and death processes (Fig 1).

Description of the process

Given a population that can contain up to m different kinds of entities, it can be completely characterized by specifying the number of individuals of each type of entity. Thus, the state of the population at a given time t is an m-dimensional vector of the form $\mathbf{n} = [n_1(t), n_2(t), \ldots, n_m(t)]^{\mathrm{T}}$, where $n_i(t)$ is the number of individuals of type i.

We assume that the birth and death rate of each type in the population depends only on the state of the population (the vector $\mathbf{n}$), and thus neglect any potential contributions from a temporally varying external environment. Our model unfolds in continuous time, and we assume that the probability of two or more births (or deaths) occurring at the same instant is negligible. For each type $i \in \{1, 2, \ldots, m\}$, we denote the birth rate and the death rate by $b_i(\mathbf{n})$ and $d_i(\mathbf{n})$, respectively. We assume that the birth and death rates at the population level scale with the total population size such that $b_i(\mathbf{n})$ and $b_i(\mathbf{n})$ are of the order of $\sum_i n_i$. Further, we assume that there exists a *carrying capacity* or more generally a *a population size measure* (Czuppon and Traulsen, 2021) $K > 0$ that imposes a bound on population growth rate such that the growth rate of the total population size $\sum_i n_i$ is expected to be negative whenever $\sum_i n_i > K$.

We can now define a notion of population density $\mathbf{x} = \mathbf{n}/K$ by dividing the population number by the carrying capacity. In terms of population densities, our assumption on the bound on growth rate translates to the growth rate of population density being negative whenever the density exceeds 1. Naturally, the limit $K \to \infty$ corresponds to the limit of infinitely large populations. Note that we may still speak of population densities in the infinite population size limit since population densities remain finite.

Functional forms of the birth and death rates

In mathematical terms, the above assumptions on population growth rates and birth and death rates amount to saying that we can find $\mathcal{O}(1)$ functions $b_i^{(K)}$ and $d_i^{(K)}$ such that we can write

$$\begin{aligned} b_i(\mathbf{n}) &= Kb_i^{(K)}(\mathbf{n}/K) \\ d_i(\mathbf{n}) &= Kd_i^{(K)}(\mathbf{n}/K) \end{aligned} \tag{1}$$

Further, we assume that the birth and death rate functions have the functional form

$$\begin{aligned} b_i^{(K)}(\mathbf{x}) &= x_i b_i^{(\text{ind})}(\mathbf{x}) + \lambda Q_i(\mathbf{x}) \\ d_i^{(K)}(\mathbf{x}) &= x_i d_i^{(\text{ind})}(\mathbf{x}) \end{aligned} \tag{2}$$

where $b_i^{(\text{ind})}(\mathbf{x})$ and $d_i^{(\text{ind})}(\mathbf{x})$ are non-negative functions that respectively describe the per-capita birth and death rate of type i individuals. Mutation or migration that is of the form $x_i f(\mathbf{x})$ for some function f can simply be subsumed into the per-capita rates $b_i^{(\text{ind})}(\mathbf{x})$ and $d_i^{(\text{ind})}(\mathbf{x})$. However, the birth rate of type i individuals may contain a component that does not depend purely multiplicatively on the density of type i: For example, when $x_i = 0$, *i.e.* there are no type i individuals in the population, individuals of type i may still be born through mutations of other types or immigration from other sources (gene flow). To account for this possibility, we include the second term λQ_i in the birth rate function, as we explain in detail below. Note that no analogous problem exists for the death rate, since the death rate of type i individuals must be 0 when x_i is 0 to ensure that we never have negative population densities.

The term λQ_i in Eq. 2 models an influx of type i individuals from sources other than the existing pool of type i individuals. Here, $\lambda \geq 0$ is a constant describing the rate of influx of type i individuals from sources other than the exsting pool of type i individuals, and $Q_i(\mathbf{x})$ is a non-negative function that describes this additional contribution. For example, if type i individuals can arise due to mutations of offspring of other types of individuals during birth, λ would represent a mutation rate (typically denoted by μ) and Q_i would model the functional form of mutation. A common choice, for example, is $Q_i(\mathbf{x}) = \sum_{j \neq i} x_j$ (*i.e.* the mutation $j \to i$ occurs at a total rate of μx_j). The influx term could also model immigration of type i individuals from other populations, since such immigration would depend not on the density of individuals x_i in our focal population, but on the density of individuals in the 'source population' from which individuals are emigrating into our focal population. In this latter case, λ would represent a dispersal rate and Q_i would model the dispersal.

Our assumptions of the functional forms given by Eq. 2 thus amount to saying that birth and death rates of type i are in a form that allows us to write down per-capita birth and death rates of type i individuals, except for a potential extra influx term $Q_i(\mathbf{x})$ whose strength is controlled by an influx rate λ (assumed the same for all types). We emphasize that these birth and death rates can incorporate complicated interactions, but as we will see, the particular forms of these rate functions do not matter for our purposes as long as certain basic mathematical scaling assumptions are met (see Supplementary section S1).

We define the *Malthusian fitness* of the i^{th} type as

$$w_i(\mathbf{x}) := b_i^{(\text{ind})}(\mathbf{x}) - d_i^{(\text{ind})}(\mathbf{x}) \tag{3}$$

and the *per-capita turnover rate* of the i^{th} type as

$$\tau_i(\mathbf{x}) = b_i^{(\text{ind})}(\mathbf{x}) + d_i^{(\text{ind})}(\mathbf{x}) \tag{4}$$

The quantity $w_i(\mathbf{x})$ describes the per-capita growth rate of type i individuals in a population $\mathbf{x}$, discounting the extra influx term. It is notable that both w_i and τ_i depend on the state of the population as a whole (*i.e.* $\mathbf{x}$) and not just on the density of the focal type. Thus, in general, both the fitness and the turnover rate in our model may be both density-dependent and frequency-dependent.

Statistical measures for population-level quantities

Given any state $\mathbf{x}(t)$ that describes our population at time t, let us first define the total (scaled) population size ($N_K(t)$) and the frequency $p_i(t)$ of each type i in the population at time t as:

$$\begin{aligned} N_K(t) &:= \sum_{i=1}^{m} x_i(t) = \frac{1}{K}\sum_{i=1}^{m} n_i(t) \\ p_i(t) &:= \frac{n_i(t)}{\sum\limits_{j=1}^{m} n_i(t)} = \frac{x_i(t)}{\sum\limits_{j=1}^{m} x_i(t)} = \frac{x_i(t)}{N_K(t)} \end{aligned} \tag{5}$$

N_K here is an $\mathcal{O}(1)$ quantity and KN_K is the total population size, which is $\mathcal{O}(K)$. Since N_K is the sum of m stochastically fluctuating quantities, the total population size KN_K also experiences stochastic fluctuations and is thus non-constant in our model. We use the term 'fluctuating populations' henceforth to refer to populations of non-constant size that experience stochastic fluctuations in this manner.

Note that while it may appear as if we have increased the number of dimensions of the system by 1, the frequency vector is subject to the constraint $\sum_i p_i = 1$, and we thus only need to study the system using the m variables $[p_1, p_2, \ldots, p_{m-1}, N_K]$. We are often interested in tracking the effects of evolution on quantities described at a population level. To facilitate this, let f be any quantity that can be defined at the type-level, such as phenotype or fitness, with a (possibly time-dependent) value $f_i \in \mathbb{R}$ for the i^{th} type. Recall that we defined m discrete types in the population on the basis that individuals within each type can be approximated as identical.

Now, the statistical mean value of such a quantity in the population $[p_1, p_2, \ldots, p_m]$, which we denote by $\bar{f}$, is given by

$$\overline{f}(t) := \sum_{i=1}^{m} f_i p_i \tag{6}$$

while the statistical covariance of two such quantities f and g in the population is given by

$$\mathrm{Cov}(f,g) := \overline{fg} - \overline{f}\,\overline{g} \tag{7}$$

Lastly, the statistical variance of a quantity f in the population is given by $\sigma_f^2 := \mathrm{Cov}(f,f)$. It is important to recognize that these statistical quantities are distinct from and independent of the *probabilistic* expectation, variance, and covariance obtained by integrating over realizations in the underlying probability space. We will denote this latter expectation and variance by $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ respectively for clarity.

Fundamental equations of eco-evolutionary dynamics

Ecological dynamics: Changes in population density

Having defined key assumptions of our eco-evolutionary dynamics via a generic birth and death process, we now proceed to understand how the population density vector $\mathbf{x}$, changes over time.

Recall that the stochastic birth-death process changes in units of $1/K$ in density space. Thus, if K is large, each individual contributes a negligible amount to the population density, and the discontinuous jumps due to individual-level births or deaths in units of $1/K$ can be approximated as small, *continuous* changes in population density $\mathbf{x}$. In Supplementary section S1, we use a formal version of this intuitive idea via a 'system size expansion' (Ethier and Kurtz, 1986, Chapter 11; Van Kampen, 1981, Chapter 10; Black and McKane, 2012; Czuppon and Traulsen, 2021) to derive a continuous description of the stochastic process for population densities. This continuous description takes the form of an Itô stochastic differential equation (Itô SDE) which says that the density of the i^{th} type changes according to

$$dx_i = \left[x_i w_i(\mathbf{x}) + \lambda Q_i(\mathbf{x})\right] dt + \frac{1}{\sqrt{K}} \sqrt{x_i \tau_i(\mathbf{x}) + \lambda Q_i(\mathbf{x})}\, dW_t^{(i)} \tag{8}$$

where each $W_t^{(i)}$ is an independent one-dimensional Wiener process (standard Brownian motion); recall that w_i and τ_i are Malthusian fitness and turnover rate of type i, respectively, as defined in Equations (3) and (4) whereas λQ_i captures the influx.

The first and second terms on the RHS of Eq. 8 respectively provide the so-called 'infinitesimal mean' and 'infinitesimal variance' of the stochastic process $x_i(t)$ that satisfies Eq. 8 (Karlin and Taylor, 1981; Czuppon and Traulsen, 2021). Informally, the infinitesimal mean and variance can be understood as follows: If we imagine that the population density of type i changes from x_i to $x_i + dx_i$ over a very small (infinitesimal) time interval dt, we can (informally) view dx_i as a random variable. In that case, the *expected* density change $\mathbb{E}[dx_i]$ and the *variance* in the change $\mathbb{V}[dx_i]$ are respectively given by:

$$\mathbb{E}[dx_i] = x_i w_i(\mathbf{x}) + \lambda Q_i(\mathbf{x}) \tag{9a}$$

$$\mathbb{V}[dx_i] = \frac{1}{K}\left(x_i \tau_i(\mathbf{x}) + \lambda Q_i(\mathbf{x})\right) \tag{9b}$$

Thus, the Malthusian fitness w_i controls the expected change in population density, whereas the turnover rate τ_i (which is also a measure of the total number of events experienced by type i in a given time interval) controls the variance in the change in population density. Note that unlike in classic bet-hedging (Gillespie, 1974), an additional influx of individuals as modeled by λQ_i could mean that τ_i is *not* equal to the infinitesimal variance, but is simply proportional to it.

Eq. 8 describes the ecological population dynamics. However, evolution is described not in terms of population density, but in terms of trait frequencies. Thus, to study the evolutionary dynamics of finite populations, we need to move from population densities $\mathbf{x}$ to trait frequencies $\mathbf{p}$. This seemingly innocuous observation has important consequences for evolutionary dynamics, as we explain below and in Box 1.

Replicator-mutator equation for finite fluctuating populations

We now use Itô calculus to derive equations for the evolutionary dynamics of trait frequencies from Eq. 8, our SDE for population densities. Letting $\overline{w} = \sum w_i p_i$ and $\overline{\tau} = \sum \tau_i p_i$ be the average population fitness and the average population turnover respectively, we show in Supplementary section S2 that p_i, the frequency of the i^{th} type in the population $\mathbf{x}(t)$, changes according to the equation:

$$\begin{aligned} dp_i(t) = {} & \underbrace{\left[\left(w_i(\mathbf{p}, N_K) - \overline{w}\right)p_i\right] dt}_{\substack{\text{Natural Selection} \\ \text{(higher fitness is better)}}} - \underbrace{\frac{1}{KN_K(t)}\left[\left(\tau_i(\mathbf{p}, N_K) - \overline{\tau}\right)p_i\right] dt}_{\substack{\text{Noise-induced selection via} \\ \text{Gillespie effect (lower turnover is better)}}} \\ & \underbrace{+ \lambda\left(1 - \frac{1}{KN_K(t)}\right)\left\{Q_i(\mathbf{p}) - p_i\left(\sum_{j=1}^{m} Q_j(\mathbf{p})\right)\right\} dt}_{\substack{\text{Transmission effects due to} \\ \text{influx of individuals from other sources} \\ \text{(mutation, migration, etc.)}}} + \underbrace{\frac{1}{\sqrt{KN_K(t)}} dW_p}_{\substack{\text{Stochastic fluctuations} \\ \text{(Non-directional} \\ \text{over small timescales)}}} \end{aligned} \tag{10}$$

where we have defined $Q_i(\mathbf{p}) := Q_i(\mathbf{x})/N_K(t)$. Here, W_p is a stochastic integral term given by

$$dW_p := \sqrt{p_i(1-p_i)^2\tau_i + p_i^2\left(\sum_{j\neq i}\tau_j p_j\right) + \lambda\left((1-p_i)^2 Q_i(\mathbf{p}) + p_i^2\sum_{j\neq i} Q_j(\mathbf{p})\right)}\, dW_t \tag{11}$$

where W_t is a one-dimensional Wiener process. The first term of Eq. 10 represents the effect of natural selection for increased (Malthusian) fitness. Eq. 10 recovers the replicator-mutator equation in the infinite population ($K \to \infty$) limit when Q_i models mutation and λ is a mutation rate (see section S7 in the supplementary). However, finite populations experience a directional force dependent on $\tau_i(\mathbf{x})$, the per-capita turnover rate of type i, that cannot be captured in infinite population models (Week et al., 2021; Kuosmanen et al., 2022). This term shows that the effect of turnover rates is structurally identical to that of the differential fitness, but it acts in the opposite direction - a higher relative τ_i leads to a decrease in frequency (Notice the minus sign before the second term in Eq. 10). For this reason, the effect has been termed 'noise-induced selection' (Week et al., 2021), though the same general idea has also been known under the names 'bet-hedging' and 'Gillespie effect' in the evolutionary ecology literature (Gillespie, 1974; Gillespie, 1977; Frank and Slatkin, 1990; Starrfelt and Kokko, 2012; Veller et al., 2017). Noise-induced selection acting through the second term of the RHS of Eq. 10 can be heuristically understood as a stochastic selection for reduced variance in changes in population density (Box 1).

The third term on the RHS of Eq. 10 represents potential biasing effects due to the influx of individuals of type i in a manner that does not depend purely multiplicatively on the current population density x_i of type i individuals (for example, through immigration from an external population or mutation of other types during birth). Since $1 - 1/KN_K$ is typically very close to 1 for medium to large population size (KN_K), we see that such influxes of individuals are not strongly affected by demographic stochasticity and thus have qualitatively similar effects in small, large, and infinite populations.

Finally, the last term describes the effects of stochastic fluctuations due to the finite size of the population and shows the $1/\sqrt{KN_K}$ scaling that is typical of demographic stochasticity. Though this last term vanishes upon taking probabilistic expectations (and is hence 'non-directional' in the short term), it may have important consequences on long-term evolutionary trajectories, as we illustrate in the next section.

To complete the description of the system, we also require an equation for the total scaled population size $N_K = \sum x_i$. Upon noting that $dN_K = \sum dx_i$ and using Eq. 8 for dx_i, and upon dividing both sides by N_K we find

$$\frac{1}{N_K}dN_K = \left[\overline{w}(t) + \lambda\sum_{i=1}^{m} Q_i(\mathbf{p})\right]dt + \frac{1}{\sqrt{KN_K(t)}}\left[\overline{\tau}(t) + \lambda\sum_{i=1}^{m} Q_i(\mathbf{p})\right]^{1/2} dW_t^{N_K} \tag{12}$$

where $W_t^{N_K}$ is a one-dimensional Wiener process and we have used the representation of noise terms presented in Supplementary section S5. Note that fitness affects only the deterministic term, turnover rate affects only the stochastic term, and potential influxes of individuals due to mutations at birth contribute to both the deterministic and the stochastic terms of Eq. 12.

Eq. 10 and 12 together completely specify the system. Since the influx terms λQ_i are minimally affected by demographic stochasticity in Eq. 10 and contribute in the same way to both terms on the RHS of Eq. 12, we do not expect this term to cause major qualitative differences in the evolutionary dynamics of finite vs infinite populations. Keeping the goal of conceptual synthesis and clarity in mind, we will therefore omit these terms in all subsequent equations presented in the main text by setting $\lambda = 0$. However, we carry out all derivations with the influx terms intact in the Supplementary sections (S1-S4), which is where we refer the interested reader.

A special case: Two interacting types

To illustrate the way stochasticity affects evolutionary dynamics in finite, fluctuating populations, we consider the simple case of two interacting types with no additional influx terms in either type (i.e. $m = 2, \lambda = 0$). Letting $p = p_1$ be the frequency of type 1 individuals in the population, we find that in (p, N_K) space, our system obeys the equations

$$\begin{aligned} dp = &\left[(w_1 - w_2)p(1-p) - \frac{1}{KN_K}(\tau_1 - \tau_2)p(1-p)\right]dt \\ &+ \frac{1}{\sqrt{KN_K}}\sqrt{p(1-p)\left[\tau_1 + (\tau_2 - \tau_1)p\right]}dW_t \end{aligned} \tag{13a}$$

$$\frac{1}{N_K}dN_K = \overline{w}(t)dt + \sqrt{\frac{\overline{\tau}(t)}{KN_K(t)}}dW_t^{N_K} \tag{13b}$$

where W_t and $W_t^{N_K}$ are one-dimensional Wiener processes. We can now identify the (frequency-dependent) selection coefficient $s(p, N_K) := w_1(p, N_K) - w_2(p, N_K)$ from classic population genetics. The selection coefficient quantifies the direction and strength of natural selection in the system — a positive value of s indicates that type 1 individuals are favored by natural selection, and a negative value of s indicates that type 1 individuals are disfavored by natural selection.

Eq. 13a also motivates the definition of an analogous noise-induced selection coefficient $\kappa(p, N_K) := \tau_2(p, N_K) - \tau_1(p, N_K)$ to quantify the direction and strength of noise-induced selection. If type 1 has a lower turnover rate, $\kappa(p, N_K) > 0$, and thus type 1 is favored by noise-induced selection.

With this notation, Eq. 13a becomes

$$dp = p(1-p)\left[s(p, N_K) + \frac{\kappa(p, N_K)}{KN_K}\right]dt + \frac{1}{\sqrt{KN_K}}\sqrt{p(1-p)\left(\tau_1 + p\kappa(p, N_K)\right)}dW_t \tag{14}$$

Box 1: A heuristic for noise-induced selection over small time intervals

One key mechanism through which noise-induced selection can affect evolutionary dynamics is by biasing evolutionary trajectories towards types with lower turnover rates, even if these types have the same (or even lower) fitness than other types in the population. Here, we explain this mechanism via an intuitive argument that has the same flavor as arguments seen in the bet-hedging literature (Gillespie, 1977; Frank and Slatkin, 1990; Starrfelt and Kokko, 2012). Ignoring influx terms ($\lambda = 0$), Eq. 8 becomes

$$dx_i = x_i w_i(\mathbf{x})dt + \sqrt{\frac{x_i \tau_i(\mathbf{x})}{K}}dW_t^{(i)} \tag{i}$$

To illustrate the idea via an example, imagine a system consisting of two types of individuals, 1 and 2, which have equal fitness but unequal turnover rates; without loss of generality, assume $\tau_1 > \tau_2$. Let us further assume that both types have the same density x_0. From Eq. (9), we see that the infinitesimal mean of population density is given by $x_i w_i(\mathbf{x})$, whereas the infinitesimal variance is given by $x_i \tau_i(\mathbf{x})/K$. Thus, in our example, though the two types of individuals have the *same expected change* in population density, type 1 individuals have a *greater variance* in the changes in density than type 2 individuals.

Since evolution is defined as changes in trait frequencies, we transform variables from population density to trait frequency to see how differential variance affects evolutionary trajectories. This is done via the transformation

$$p_i = \frac{x_i}{x_i + \sum\limits_{j \neq i} x_j} \qquad \text{for any fixed } i \in \{0, 1, 2, \dots, m\}, \tag{ii}$$

Observe now that frequency (p_i) is a *concave* function of density x_i (Eq (ii)). Due to concavity, equivalent changes in density do not correspond to equivalent changes in frequency. Instead, a result mathematically known as Jensen's inequality and diagrammatically represented in figure 2 applies.

Note that an increase in density leads to a relatively smaller increase in frequency, whereas an equivalent decrease in density leads to a larger decrease in frequency. This implies that stochastic reductions in density have a higher cost (decrease in frequency) than the benefits (increase in frequency) conferred by a numerically equivalent increase in density (Fig. 2). Thus, variance in the density process leads to a net cost in frequency space, and all else being equal, a greater variance comes with a greater cost. Types with lower turnover rates (corresponding to lower infinitesimal variance in Eq. (i)) are thus favored.

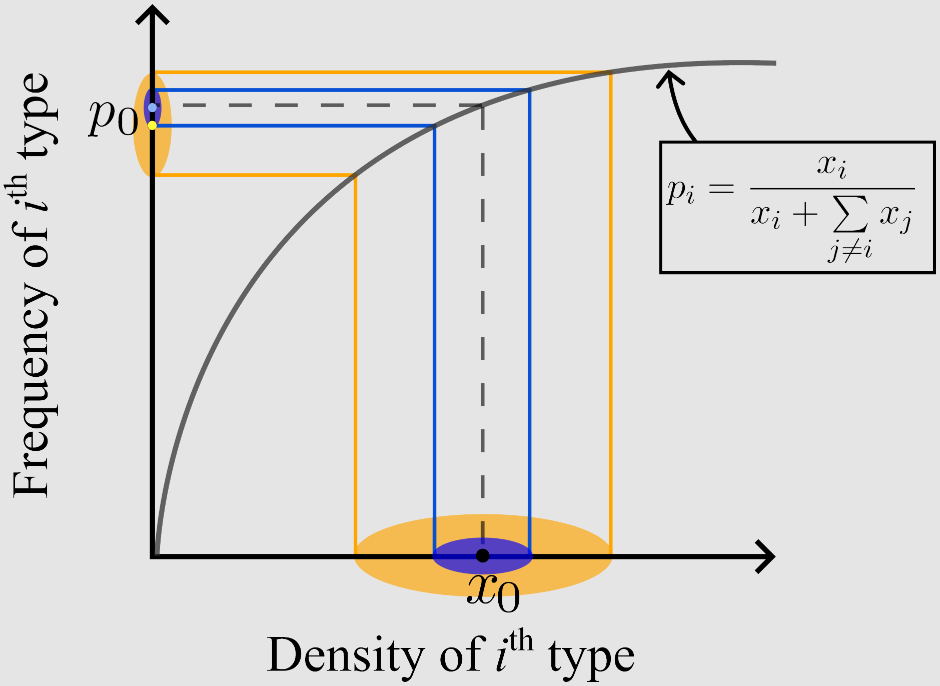

Figure 2: A diagrammatic representation of the consequences of demographic stochasticity when total population size can vary. The grey curve represents the transformation from population densities to trait frequencies via Eq. (ii). The ellipses are representations of possible changes in population composition for two types with the same fitness and same initial density, but different variances (yellow $>$ blue). The center of the ellipse represents the infinitesimal mean of the density process, the major axis captures the infinitesimal variance, and the colored region is thus representative of all possible changes given that an event (birth or death) has occurred. Reductions in density have a stronger effect on frequency than increases in density, and due to this, the expected frequency (centers of ellipses on the y-axis) after an event has occurred is less than the initial frequency p_0 even if the expected density (centers of ellipses on the x-axis) coincides with the initial density x_0. Types with a larger variance in the density process (yellow ellipse in the figure) experience a greater reduction in expected frequency relative to types with a lower variance (blue ellipse). Similar figures, with the X and Y axes being absolute fitness and relative fitness respectively, appear in expositions of bet-hedging (e.g. Frank and Slatkin, 1990; Starrfelt and Kokko, 2012); In our figure, the axes are population density and trait frequency respectively.

The argument we provide here is particular to populations of non-constant size. To see this, assume that the total (scaled) population size $\sum_i x_i$ is a constant $N > 0$. The transformation in Eq. (ii) then becomes

$$p_i = \frac{x_i}{x_i + \sum\limits_{j \neq i} x_j} = \frac{x_i}{N} \tag{iii}$$

and is now simply a linear re-scaling of x_i rather than a concave function. The asymmetry between increases in density and decreases in density observed in Fig. 2 thus disappears. In other words, the mechanism that we identified above no longer works! 321

where we see that the selection coefficient $s(p, N_K)$ affects the dt term of Eq. 14, and the noise-induced selection coefficient $\kappa(p, N_K)$ affects both the dt and dW_t terms. Note that fitness only enters into the population dynamics via the selection coefficient s, whereas turnover also appears via τ_1 in the second term on the RHS of Eq. 14. In other words, only the relative fitness or the difference $w_1 - w_2$, but not the absolute value of the fitness w_i, matters for the deterministic dynamics. In contrast, the absolute value of the per-capita turnover rate *does* affect the stochastic dynamics of the system via the second term on the RHS of Eq. 14.

Noise-induced selection can also affect the long-term behaviour of the population dynamics through the second term on the RHS of Eq. 14 due to turnover-dependent stochastic effects (McLeod and Day, 2019a). Let $m(p)$ be the probability density function associated with the quasi-stationary distribution of the stochastic process defined by Eq. 14. Informally, if we restrict ourselves to situations in which neither type has reached fixation or gone extinct in the population, $m(p)dp$ describes (upto a normalization constant) the probability of observing a trait frequency in the interval $(p, p + dp)$ for a very small value of dp. In supplementary section S6, we show (Eq. S80) that the quasi-stationary density $m(p)$ obeys the equation:

$$\frac{dm}{dp} = m(p)\left[\underbrace{\frac{2p-1}{p(1-p)}}_{\substack{\text{Anti-symmetric}\\ \text{about } p=0.5}} + \underbrace{2\frac{E(p)}{V(p)}}_{\substack{\text{Same sign}\\ \text{as first term}\\ \text{on RHS of Eq.14}}} - \underbrace{\frac{1}{V(p)}\frac{dV}{dp}}_{\substack{\text{Contributions from}\\ \text{second term}\\ \text{on RHS of Eq. 14}}} \right] \tag{15}$$

where

$$E(p, N_K) = s(p, N_K) + \frac{1}{KN_K}\kappa(p, N_K) \tag{16a}$$

$$V(p, N_K) = \frac{1}{KN_K}\left(\tau_1(p, N_K) + p\kappa(p, N_K)\right) \tag{16b}$$

The sign of dm/dp tells us whether type 1 is favored in the stationary distribution (positive meaning that type 1 is favored), and points at which $dm/dp = 0$ tell us about the most likely and/or least likely value of p in the stationary distribution (McLeod and Day, 2019a; Majumder et al., 2021). The first term on the RHS of 15 is anti-symmetric about $p = 0.5$ and thus uninteresting on its own for determining the sign of dm/dp as a function of p.

The second term of Eq. 15 represents the balance between classical selection and a form of noise-induced selection that is visible over short timescales (see Box 1). Since both s and κ are $\mathcal{O}(1)$ functions, natural selection will tend to dominate $E(p)$ when the total population size KN_K is large. Additionally, if s and κ are of similar magnitude (*i.e.* the strength of natural selection is comparable to the strength of the Gillespie effect), natural selection will still dominate the sign of $E(p)$ since the total population size KN_K must be greater than 1. However, noise-induced selection in this form can qualitatively affect evolutionary dynamics if differences in Malthusian fitnesses are close to zero (*i.e.* natural selection is weak, $s \approx 0$) or if total population size KN_K is small. We will also show this explicitly using an example in the next section.

Eq. 15 also tells us that the 'Gillespie effect' explained in Box 1 is not the only way in which noise-induced selection can affect evolutionary dynamics over long timescales: Instead, the long-term behaviour of finite population systems is also profoundly affected by the 'noise' terms in Eq. 14, as captured by the last term on the RHS of Eq. 15. In particular, even when the first term on the RHS of Eq. 14 vanishes or acts in the same direction as classical selection (*i.e.* the 'Gillespie effect' is weak or absent), the long-term behavior of finite populations may still systematically differ from infinite population predictions, in particular possibly 'reversing' the direction of evolution, if dV/dp is non-zero. For example, type 1 individuals can be overrepresented at equilibrium even if $s + \kappa/KN_K < 0$ (meaning that the first term on the RHS of 14 favors type 2 individuals) as long as dV/dp is sufficiently negative (McLeod and Day, 2019a). As an aside, note that $1/VdV/dP$ could also equivalently be written as the derivative of $\log(V)$ with respect to p, and thus represents the strength and direction of frequency dependence of $\log(V)$. Since $\mathbb{V}[dp] = p(1-p)V(p)$ from Eq. 14, $\log(V)$ can be interpreted as being proportional to the logarithm of the variance in the changes in the trait frequency dp. This term thus captures the contributions of stochastic fluctuations/'noise' in the trait frequency changes dp and can be interpreted as 'selecting' for reduced variance in the change in trait frequencies dp, whereas the 'Gillespie effect' is a selection for reduced variance in the change in population densities. To the best of our knowledge, this effect was first explicitly recognized in the literature by McLeod and Day (2019a) in the context of social evolution.

Remarkably, in the case where natural selection does not operate ($s = 0$), if τ_1 and κ are independent of the total population size KN_K, so is the stationary distribution. This can be seen by noting that when $s = 0$, the total population size KN_K affects the dynamics only through a pre-factor of $1/KN_K$ that occurs in both $E(p)$ and $V(p)$. It therefore disappears in the ratio E/V. Thus, unlike the classic results regarding the relative strengths of natural selection and genetic drift, the total population size does *not* affect the relative strengths of noise-induced selection and genetic drift — instead, it is the difference in turnover rates κ and the frequency-dependence of $V(p)$ that determine whether the stationary distribution favors one type over the other. A similar observation has been made in the context of life-history theory (Shpak, 2005).

Noise-induced selection clearly disappears in the infinite population limit ($K \to \infty$). Additionally, we can now make our claim from Box 1 that noise-induced selection is particular to populations of non-constant total population size precise: If the total population size $KN_K = K\sum_j x_j$ is a constant, the additional stochastic term introduced by Itô's formula in the derivation carried out in S2 vanishes. Instead, simply dividing the equation for species densities (Eq. 8) by the (now constant) total population size directly provides the complete dynamics of the system in frequency space. The deterministic part of Eq. 8 depends only on the fitness w_i and influx terms but does not depend on turnover rate τ_i. We can therefore also conclude that noise-induced selection through the Gillespie effect is particular to *non-constant* populations whose total population size KN_K changes stochastically depending on the population composition $\mathbf{x}(t)$.

An example with two competing types

To illustrate when noise-induced selection can be important for population dynamics, we use a simple biologically motivated example in this section. Several abiotic factors such as temperature and pH are known to be ecological 'rate modulators' that affect either the birth rate or death rate of organisms, with obvious consequences for evolutionary dynamics (Fronhofer et al., 2023). To see how demographic stochasticity may affect the effect of ecological rate modulators on evolutionary dynamics, consider here two competing phenotypes, which we denote by 1 and 2. Though we stick to this 'rate modulation' language henceforth, another potential interpretation of the model we study below comes from epidemiology: In this case, the two types can be thought of as two competing strains of pathogens, a 'rate modulator' that affects birth rates can be thought of as affecting transmission rate, and a 'rate modulator' that affects death rates can be thought of as affecting virulence (Parsons et al., 2018). We consider the case where type 1 is affected by the ecological rate modulator but type 2 is not. For simplicity, we assume the population is closed with no mutations during birth (*i.e.* $\lambda = 0$). Below, we use p to denote the frequency of type 1 individuals in the population.

For pedagogical clarity, we assume that rate modulation occurs by simply shifting the birth and/or death rate by a constant. In equations, this can be modelled via the relations:

$$b_1^{(\mathrm{ind})}(p, N_K) = b_2^{(\mathrm{ind})}(p, N_K) + \epsilon_b \tag{17a}$$

$$d_1^{(\mathrm{ind})}(p, N_K) = d_2^{(\mathrm{ind})}(p, N_K) + \epsilon_d \tag{17b}$$

where ϵ_b and ϵ_d are real numbers describing the effect of the ecological rate modulator on the birth and death rates respectively. Using the definitions of s and κ, we find

$$s(p, N_K) = \epsilon_b - \epsilon_d \tag{18a}$$

$$\kappa(p, N_K) = -[\epsilon_b + \epsilon_d] \tag{18b}$$

Note that if $\epsilon_b = 0, \epsilon_d < 0$, both s and κ are positive, whereas if $\epsilon_b > 0, \epsilon_d = 0$, $s > 0$ but $\kappa < 0$. In other words, if type 1 has a decreased death rate (virulence in the epidemiological case) but identical birth rate relative to type 2, type 1 is favored by both natural selection and noise-induced selection. On the other hand, if type 1 has an increased birth rate (transmission rate in the epidemiology case) but an identical death rate relative to type 2, type 1 is favored by natural selection but disfavored by noise-induced selection. Thus, all else being equal, reducing the death rate is generically more favorable than increasing the birth rate by an analogous amount, an observation that has been made in finite population models in epidemiology (Parsons et al., 2018), social evolution (McLeod and Day, 2019a), life-history evolution (Alexander and Wahl, 2008), and cancer biology (Raatz and Traulsen, 2023).

For the rest of this example, we assume that $\epsilon_b > 0, \epsilon_d > 0$, *i.e.* that type 1 has both an increased birth rate and an increased death rate compared to type 2. We may now ask, when is the outcome of evolution different from that expected by infinite population dynamics?

Noise-induced selection in the absence of natural selection

First, consider the situation $\epsilon_b = \epsilon_d = \epsilon$. This corresponds to the two types having the same growth rate, but type 1 having a faster pace of life than type 2. The selection coefficient and noise-induced selection coefficient are

$$s(p, N_K) = 0 \tag{19}$$

$$\kappa(p, N_K) = -2\epsilon \tag{20}$$

Thus, as expected, natural selection does not operate in the system. In the infinite population limit, natural selection is the only force that affects population dynamics and we thus expect any initial frequency p_0 of type 1 individuals to remain unchanged in the population (to see this, take $K \to \infty$ in Eq. 14). Over short timescales, the effects of demographic stochasticity can be observed by looking at the expected change in frequency $\mathbb{E}[dp]$. Using Eq. 14 and substituting the functional forms given by Eq. 18, we find

$$\frac{d}{dt}\mathbb{E}[p] = \mathbb{E}\left[\frac{\kappa(p, N_K)}{N_K}p(1-p)\right] = -\frac{2\epsilon}{KN_K}\mathbb{E}[p(1-p)] \tag{21}$$

Since the RHS of Eq. 21 is always negative for $p \in (0,1)$, we can infer that if the system begins at any initial frequency $p_0 \in (0,1)$, the proportion of type 1 individuals is expected to decrease. If $\epsilon_b = \epsilon_d$, the ecological rate modulator is thus detrimental to the evolutionary fate of type 1 individuals over short time scales in finite populations, despite infinite population models predicting neutrality. This result is a manifestation of the 'fast' mechanism of noise-induced selection via the Gillespie effect from Box 1: All else being equal, a faster pace of life comes with a greater variance in change of population density within a given time interval since there are simply more stochastic birth/death events taking place.

However, the evolutionary fate over long timescales depends not only on the expected change of frequency alone but also on the variance in the change of frequencies. This stochastic effect, captured by the dW term in Eq. 14, depends on the functional form of $\tau_1(p, N_K)$ (and not merely the difference $\kappa = \tau_2 - \tau_1$), which we have not yet specified in our model (Eq. 17). For simplicity, let us assume that the turnover rates τ_i have linear frequency dependence. Specifically, let us assume $\tau_1 = bp + c$, where p is the frequency of type 1 individuals and b and c are suitable constants. Since we would like κ to still be given by Eq. 20, this automatically specifies τ_2 as $\tau_2 = bp + c - 2\epsilon$. Thus, we assume τ_1 and τ_2 change in the same direction (increase or decrease) as the frequency of type 1 individuals increases.

The probability of observing a population in which the trait frequency is p can be described via the so-called 'quasi-stationary density' (speed measure). The quasi-stationary density describes the probability of observing the population in a given configuration (p, N_K) conditioned on non-extinction of either type (Supplementary section S6). In our example, we can derive an exact expression for the quasi-stationary distribution. We present this solution in Supplementary section S9, and only illustrate the key points here via Figure 3.

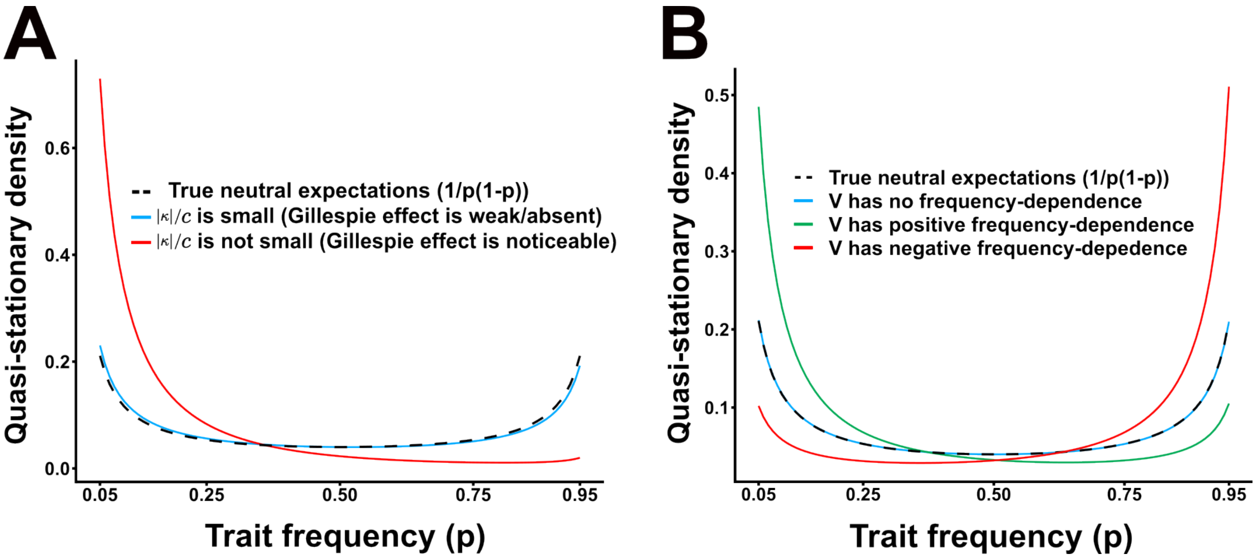

Figure 3: Two distinct noise-induced effects on evolutionary dynamics. **A.** If the magnitude of the noise-induced selection coefficient κ is large relative to the intrinsic turnover rate c, the fast mechanism of noise-induced selection for reduced per-capita variance operates. Parameters are chosen such that $V(p)$ is not frequency-dependent (blue: $\epsilon = 0.5, b = 1, c = 10$; red: $\epsilon = 0.5, b = 1, c = 0.5$). **B.** The stationary distribution can also be biased if $V(p) = \tau_1 + p\kappa = \tau_1 - 2\epsilon p$ is frequency-dependent. This slower mechanism of noise-induced selection favors the type that reduces $V(p)$. Parameters in this panel are chosen such that the strength of the fast mechanism (Gillespie effect) is negligible (blue: $\epsilon = 0.025, b = 0.05, c = 10$; green: $\epsilon = 0.025, b = 50, c = 10$; red: $\epsilon = 0.025, b = -8.5, c = 10$)

Figure 3 illustrates two distinct ways in which 'noise-induced selection' can manifest in finite, fluctuating populations. If dynamics are truly neutral (in the sense of the two types being exactly equivalent) and the system begins with $p = 0.5$, then both types are equally likely to increase/decrease. The quasi-stationary density is thus equal to $1/p(1-p)$ (up to a constant). Noise-induced effects can bias this distribution in two distinct ways (Box 2): (1) The noise-induced selection coefficient κ can bias the expected trajectory over short timescales, eventually leading to a bias in the stationary distribution. Since this effect can be observed over short timescales as deviations of dp/dt from the expected trajectory, we refer to it as the 'fast' mechanism of noise-induced selection. The fast mechanism can be identified with the Gillespie effect from bet-hedging theory and is a selection for reduced variance in density change dx_i. Since $\kappa = -2\epsilon < 0$, the fast mechanism always favors the type with the slower pace of life (Fig. 3A). (2) A second noise-induced effect appears only over long time scales through a biasing of the stationary distribution via the last term of Eq. 15. We thus call it a 'slow' mechanism of noise-induced selection. For our example, we can calculate

$$\begin{aligned} V(p) &= \tau_1 + p\kappa = bp + c + p\kappa \\ \Rightarrow \frac{dV}{dp} &= b - \kappa \end{aligned} \tag{22}$$

Equation 22 tells us that this slower mechanism favors type 1 if $b < \kappa$, type 2 if $b > \kappa$, and does not operate if $b = \kappa$. Thus, noise-induced selection through the slow mechanism may act in the same direction or the opposite direction of noise-induced selection through the fast mechanism (the 'Gillespie effect') based on the details of the frequency-dependence of the per-capita turnover rates (Fig 3B).

Noise-induced selection in the presence of natural selection

Consider now instead a situation in which the rate modulator affects the birth rate more than it does the death rate (*i.e.* $\epsilon_b > \epsilon_d > 0$). In this case, the selection coefficient s in Eq. 18 is always positive, and thus natural selection always favors type 1 individuals. As before, noise-induced selection may affect evolutionary dynamics in two distinct ways. First, noise-induced selection may invert the direction of the expected trajectory $\mathbb{E}[dp/dt]$ via the fast mechanism. Noise-induced selection may also bias the stationary distribution towards certain types through the slow mechanism of selection for reduced variance in dp. We examine the two possibilities one by one.

Since $s > 0$, we can use Eq. 14 to say the expected trajectory is in the opposite direction of infinite population predictions if $s + \kappa/KN_K < 0$. Using Eq. 18, we see that this is equivalent to

$$\begin{aligned} \epsilon_b - \epsilon_d - \frac{1}{KN_K}(\epsilon_b + \epsilon_d) < 0 &\Rightarrow \left(1 - \frac{1}{KN_K}\right)\epsilon_b < \left(1 + \frac{1}{KN_K}\right)\epsilon_d \\ &\Rightarrow \frac{\epsilon_b}{\epsilon_d} < \frac{KN_K + 1}{KN_K - 1} \end{aligned} \tag{23}$$

Using inequality 23 in Eq. 18a, we can arrive at the inequality

$$\begin{aligned} s = \epsilon_d \left(\frac{\epsilon_b}{\epsilon_d} - 1\right) &< \epsilon_d \left(\frac{KN_K + 1}{KN_K - 1} - 1\right) \\ \Rightarrow s(KN_K - 1) &< 2\epsilon_d \end{aligned} \tag{24}$$

Thus, noise-induced selection can reverse the expected trajectory of evolutionary dynamics when the product $s(KN_K - 1)$ is sufficiently small, *i.e.* when either *selection is weak* (s is small), *populations are small* (KN_K is small), or both.

We now also examine the contributions of the noise terms to the stationary distribution. We see from Eq. 15 that we can say noise-induced selection favors type 1 through the noise term when $dV/dp < 0$. Using the definition of V from Eq. 16b and substituting the functional forms given by Eq. 18, we see that $dV/dp < 0$ is equivalent to

$$\frac{d\tau_1}{dp} < \epsilon_b + \epsilon_d \tag{25}$$

If τ_1 is a constant, *i.e.* the per-capita birth rates $b_1^{(\text{ind})}$ and $d_1^{(\text{ind})}$ do not depend on population composition, inequality 25 will automatically be satisfied as long as there is some rate modulation in the system (*i.e.* ϵ_b and ϵ_d are not both 0). If τ_1 is frequency dependent, 25 is satisfied whenever τ_1 exhibits negative frequency dependence, though it may also be satisfied if τ_1 exhibits weak positive frequency dependence. We do not explore the effects of noise-induced selection on the stationary distribution further for the sake of conciseness. However, we note that since we already studied the behaviour of $E(p)$ above, it is now straightforward to determine from Eq. 15 when this latter effect combines with $E(p)/V(p)$ to make the RHS of Eq. 15 positive. In Supplementary section S10, we provide an example system in which noise-induced selection can never act through the fast mechanism to reverse the expected trajectory $\mathbb{E}[dp/dt]$, but may nevertheless affect long term evolutionary trajectories through the slow mechanism.

Box 2: Two distinct non-neutral effects of demographic stochasticity

Our equations reveal that noise-induced selection, or directional effects of demographic stochasticity, can affect evolutionary dynamics through two distinct mechanisms. In this box, we provide a synthesis of the connections and delineations between the two mechanisms.

1. The *fast mechanism* selects for reduced variance in changes in population density (Gillespie, 1974; Gillespie, 1977). Noise-induced selection, in this case, appears in the 'deterministic' term (dt term) of the replicator-mutator equation (Eq. 10) and is apparent over both short ('ecological') timescales as well as long ('evolutionary')

timescales as a systematic deviation of the expected trajectory $\mathbb{E}[dp/dt]$ from the infinite population prediction. The fast mechanism can be identified with the 'Gillespie effect' from the bet-hedging literature (Gillespie, 1974) and is obtained as a balance between natural selection for increased ecological growth rate and a stochastic selection for reduced variance in changes in population densities (see Box 1). This effect is thus a version of classical bet-hedging (Frank and Slatkin, 1990; Starrfelt and Kokko, 2012) in an explicitly demographic, dynamical context. Noise-induced selection through the fast mechanism is most apparent when natural selection is weak or absent, populations are small, or both. Non-constant total population size is essential for this effect to operate in the case of constant external environments (see Box 1).

2. The *slow mechanism* selects for reduced variance in changes in trait frequency (McLeod and Day, 2019a). Noise-induced selection in this case appears in the 'stochastic' term (dW term) of the replicator-mutator equation (Eq. 10) and is only apparent over very long ('evolutionary') timescales as a systematic bias or shift in the stationary distribution. The slow mechanism is a result of frequency-dependence in the variance of changes in trait frequencies and selects those types that are associated with lower variance (Fig 3B). The strength of slow noise-induced selection varies inversely with (the square root of) population size, and the direction of the effect depends on the frequency-dependence of the per-capita turnover rates τ_i.

Unlike natural selection, the balance between noise-induced selection (through either mechanism) and genetic drift in the absence of natural selection does *not* depend on the total population size: Instead, it is determined by the details of the demographic processes occurring in the population: If different types have different turnover rates, the fast mechanism operates, and if some types are associated with lower variance in the change in trait frequencies, the slow mechanism operates.

We provide a simple example of these effects via a model of variation in the pace of life in the main text, and the results are summarized in Fig 3. Our figure highlights that the two noise-induced effects may operate in isolation or simultaneously, and may either supplement (red curve in Fig. 3A and green curve in 3B) or oppose (red curves in 3A and 3B) each other. In Supplementary section S10, we also provide an example in which noise-induced selection can never reverse the direction of evolution predicted by natural selection through the fast mechanism but may nevertheless do so through the slow mechanism. In Supplementary section S11, we provide an example of a stochastic Lotka-Volterra competition model with both natural selection and mutation in which noise-induced selection acting via the fast mechanism can reverse the direction of evolution predicted by natural selection-mutation balance.

Price equation for finite fluctuating populations

We show that the statistical population mean $\overline{f}$ of any type-level quantity f (e.g. phenotype, fitness) changes over time according to the equation (see Supplementary section S3)

$$d\overline{f} = \underbrace{\mathrm{Cov}(w,f)dt}_{\substack{\text{Classical}\\\text{selection}}} - \underbrace{\frac{1}{KN_K(t)}\mathrm{Cov}(\tau,f)dt}_{\substack{\text{Noise-induced selection}\\\text{(fast mechanism)}}} + \underbrace{\overline{\left(\frac{\partial f}{\partial t}\right)}dt}_{\substack{\text{Ecological}\\\text{feedbacks}}} + \underbrace{\frac{1}{\sqrt{KN_K(t)}}dW_{\overline{f}}}_{\substack{\text{Stochastic}\\\text{fluctuations}}} \tag{26}$$

where

$$dW_{\overline{f}} := \left(\sqrt{\mathrm{Cov}\left(\tau,\left(f-\overline{f}\right)^2\right) + \overline{\tau}\sigma_f^2}\right)dW_t \tag{27}$$

where is a stochastic integral term describing un-directed stochastic fluctuations (see Eq. S64 in Supplementary section S5). Here, W_t is a Wiener process that is not the same Wiener process that appears in Eq. 10.

Eq. 26 recovers the Price equation in the infinite population ($K \to \infty$) limit (see section S7). Each term in Eq. 26 lends itself to a simple biological interpretation. The first term, $\mathrm{Cov}(w,f)$, is well-understood in the classical Price equation and represents the effect of natural selection. If the trait and the fitness are positively correlated, the mean trait value in the population increases due to the effect of selection. The second term, $\mathrm{Cov}(\tau,f)/KN_K(t)$ is the effect of noise-induced selection on the population mean via the fast mechanism in finite fluctuating populations. Biologically, the $\mathrm{Cov}(\tau,f)$ term (with negative sign) describes a biasing effect due to differential turnover rates between different types; if the trait is positively correlated with turnover rate, mean trait value will reduce.

The third term of Eq. 26 is relevant in both finite and infinite populations whenever f_i can vary over time and represents feedback effects on the quantity f_i of the i^{th} species over short ('ecological') time-scales. Such feedback could be through a changing environment, phenotypic/behavioral plasticity, or any manner of other 'ecological' phenomena. This is the term that captures eco-evolutionary feedback loops.

Finally, the last term of Eq. 26 describes the role of stochastic fluctuations. Recall that the square of this term corresponds to the infinitesimal variance of the change in the mean value $d\overline{f}$ of the quantity f in the population. $(f_i - \overline{f})^2$ is a measure of the distance of f_i from the population mean $\overline{f}$. The $\mathrm{Cov}\left(\tau,\left(f-\overline{f}\right)^2\right)$ term thus tells us that if turnover τ_i of the i^{th} type covaries positively with the distance of f_i from the population mean (*i.e.* individuals with more extreme f have higher turnover rates), the population experiences a greater variance in $d\overline{f}$, i.e. the change in the mean value of f over infinitesimal time intervals. The $\overline{\tau}\sigma^2$ term tells us that even if τ and f do not covary, there is still some variance in $d\overline{f}$, given now by the product of the mean turnover rate $\overline{\tau}$ with the standing variation σ_f^2 in the quantity f. As we shall see in the next section, this is a manifestation of neutral genetic/ecological drift. Just as in the replicator-mutator equation, stochastic fluctuations through $dW_{\overline{f}}$ can profoundly affect the long-term behaviour (stationary distribution) of $\overline{f}$ via the 'slow' mechanism of noise-induced selection if the term inside the square root of Eq. 27 depends on $\overline{f}$. Note that unlike for the replicator equation, the SDE in Eq. 26 is one-dimensional regardless of the number of traits (m), and thus the stationary distribution of the mean value $\overline{f}$ can always be studied the way we studied Eq. 14.

Fisher's fundamental theorem for finite fluctuating populations

Two particularly interesting implications of Eq. 26 are realized upon considering the time evolution of mean fitness and mean turnover rate. First, upon substituting $f = w$ in Eq. 26 and taking expectations over the underlying probability space, we obtain:

$$\mathbb{E}\left[\frac{d\overline{w}}{dt}\right] = \underbrace{\mathbb{E}\left[\sigma_w^2\right]}_{\substack{\text{Fisher's}\\ \text{fundamental}\\ \text{theorem}}} - \underbrace{\mathbb{E}\left[\frac{\sigma^2_{b^{(\mathrm{ind})}} - \sigma^2_{d^{(\mathrm{ind})}}}{KN_K(t)}\right]}_{\substack{\text{Noise-induced}\\ \text{selection}}} + \underbrace{\mathbb{E}\left[\overline{\frac{\partial w}{\partial t}}\right]}_{\substack{\text{Eco-evolutionary}\\ \text{feedbacks to fitness}}} \tag{28}$$

The first term, σ_w^2, is the subject of Fisher's fundamental theorem (Frank and Slatkin, 1992; Kokko, 2021), and says that in infinite populations, the rate of change of mean fitness in the population is proportional only to the standing variation in fitness σ_w^2 if fitness at the type level (w_i) does not change over time. The second term of Eq. 28 is a manifestation of noise-induced selection acting via the fast mechanism and is particular to finite populations. Finally, the last term arises in both finite and infinite populations whenever w_i can vary over time (Kokko, 2021), be it through frequency-dependent selection, phenotypic plasticity, varying environments, or other ecological mechanisms, and represents feedback effects on the fitness w_i of the i^{th} species over short ('ecological') time-scales. Eq. 28 recovers the standard version of Fisher's fundamental theorem in the infinite population ($K \to \infty$) limit (see section S7).

The demographic origins of fitness differences induce quantitative corrections to Fisher's fundamental theorem in finite populations

Since $w = b^{(\mathrm{ind})} - d^{(\mathrm{ind})}$ by definition, Eq. 28 can alternatively also be written as

$$\mathbb{E}\left[\frac{d\overline{w}}{dt}\right] = \underbrace{\mathbb{E}\left[\left(1 - \frac{1}{KN_K}\right)\sigma^2_{b^{(\mathrm{ind})}}\right]}_{\substack{\text{Changes in mean fitness}\\ \text{due to differential birth rates}}} + \underbrace{\mathbb{E}\left[\left(1 + \frac{1}{KN_K}\right)\sigma^2_{d^{(\mathrm{ind})}}\right]}_{\substack{\text{Changes in mean fitness}\\ \text{due to differential death rates}}} + \underbrace{\mathbb{E}\left[\overline{\frac{\partial w}{\partial t}}\right]}_{\substack{\text{Eco-evolutionary}\\ \text{feedbacks to fitness}}} \tag{29}$$

Eq. 29 redescribes variation in fitness in terms of the more fundamental processes of birth and death. Eq. 29 also tells us that variation in death rates leads to a slightly greater rate of increase in mean fitness than an equivalent variation in birth rates. For example, if individuals differ in birth rates alone (*i.e.* $\sigma^2_{d(\text{ind})} = 0, \sigma^2_w = \sigma^2_{b(\text{ind})}$), Eq. 29 predicts that the rate of mean fitness in the absence of eco-evolutionary effects is given by $\mathbb{E}[(1 - 1/KN_K)\sigma^2_w]$. On the other hand, if individuals instead differ in death rates alone, (*i.e.* $\sigma^2_{b(\text{ind})} = 0, \sigma^2_w = \sigma^2_{d(\text{ind})}$), the rate of change of mean fitness in the absence of eco-evolutionary effects is given by $\mathbb{E}[(1 + 1/KN_K)\sigma^2_w]$, which is a slightly faster rate of convergence. Note, however, that these are only minor quantitative corrections to Fisher's fundamental theorem and the two cases exhibit the same qualitative behaviour.

An analog of Fisher's fundamental theorem for the mean turnover rate of the population

Carrying out the same steps in deriving Eq. 28 with $f = \tau$ in Eq. 26 yields a dynamical equation for the evolution of mean turnover rates and reads

$$\mathbb{E}\left[\frac{d\overline{\tau}}{dt}\right] = \underbrace{\mathbb{E}\left[\sigma^2_{b(\text{ind})} - \sigma^2_{d(\text{ind})}\right]}_{\text{Classical selection effects}} - \underbrace{\mathbb{E}\left[\frac{\sigma^2_\tau}{KN_K(t)}\right]}_{\text{Noise-induced selection effects}} + \underbrace{\mathbb{E}\left[\overline{\frac{\partial \tau}{\partial t}}\right]}_{\text{Eco-evolutionary feedbacks to } \tau_i} \tag{30}$$

This equation has been derived by Kuosmanen et al., 2022 using an alternative mathematical approximation scheme. The effect of classical natural selection on the evolution of mean turnover depends on whether birth rates or death rates have greater variation in the population, as captured by the first term — if birth rates have greater variance than death rates, then mean turnover increases due to natural selection, whereas if birth rates have lesser variance than death rates, mean turnover decreases due to natural selection. The second term of Eq. 30 appears only in finite populations and is exactly analogous to the σ^2_w term that appears in Fisher's fundamental theorem. This term says that noise-induced selection always reduces mean turnover in the population, with the rate of reduction of the mean turnover rate being proportional to the standing variation in turnover rates σ^2_τ. Finally, the last term on the RHS of Eq. 30 quantifies the effect of eco-evolutionary feedback via changes in the turnover of each type over time.

The fundamental equation for the population variance via a generalization of an equation for variance of type-level quantities

Eq. 26 is a general equation for the mean value of an arbitrary type level quantity f in the population. In many real-life situations, we are interested in not just the population mean, but also the variance of a quantity in the population. In Supplementary section S4, we show that the statistical variance of any type level quantity f obeys

$$d\sigma_f^2 = \underbrace{\text{Cov}\left(w,(f-\overline{f})^2\right)dt}_{\substack{\text{Classical}\\\text{selection}}} - \underbrace{\frac{2}{KN_K}\text{Cov}\left(\tau,(f-\overline{f})^2\right)dt}_{\substack{\text{Noise-induced selection}\\\text{(fast mechanism)}}} - \underbrace{\frac{1}{KN_K}\overline{\tau}\sigma_f^2 dt}_{\substack{\text{Genetic/Ecologial}\\\text{drift}}}$$
$$+ \underbrace{2\text{Cov}\left(\frac{\partial f}{\partial t},f\right)dt}_{\substack{\text{Ecological}\\\text{feedbacks}}} + \underbrace{\frac{1}{\sqrt{KN_K(t)}}dW_{\sigma_f^2}}_{\substack{\text{Stochastic}\\\text{fluctuations}}} \tag{31}$$

where

$$dW_{\sigma_f^2} := \sqrt{\text{Cov}\left(\tau,\left(f-\overline{f}\right)^4\right) + \overline{\tau}(\sigma_f^2)^2}\,dW_t \tag{32}$$

is a stochastic integral term measuring the (non-directional) effect of stochastic fluctuations that vanishes upon taking an expectation over the probability space (see Eq. S66 in Supplementary section S5). As before, we use W_t to denote a generic Wiener process — the W_t that appears in Eq. 31 is not necessarily the same process that appears in either Eq. 10 or Eq. 26. The stochastic dependencies between the various Wiener processes can be studied using a relation discussed in Supplementary section S5.

Once again, terms of Eq. 31 lend themselves to straightforward biological interpretation. The quantity $(f_i - \overline{f})^2$ is a measure of the distance of f_i from the population mean value $\overline{f}$, and thus covariance with $(f-\overline{f})^2$ quantifies the type of selection operating in the system: A negative correlation is indicative of stabilizing or directional selection, and a positive correlation is indicative of disruptive (*i.e.* diversifying) selection (Rice, 2004, Chapter 6; Lion, 2018). An extreme case of diversifying selection for fitness occurs if the mean fitness of the population is at a local minimum but $f_i \not\equiv \overline{f}$ (*i.e.* the population still exhibits some variation in f). In this case, if the variation in f is associated with a variation in fitness, then $\text{Cov}(w,(f-\overline{f})^2)$ is strongly positive and the population experiences a sudden explosion in variance, causing the emergence of polymorphism in the population. If $\text{Cov}(w,(f-\overline{f})^2)$ is still positive after the initial emergence of multiple morphs, evolution eventually leads to the emergence of stable coexisting polymorphisms in the population - this phenomenon is a slight generalization of the idea of evolutionary branching that occurs in frameworks such as adaptive dynamics (Doebeli, 2011). The $\text{Cov}\left(\partial f/\partial t, f\right)$ term represents the effect of eco-evolutionary feedback loops due to changes in f at the type level.

Finally, the last term on the RHS of Eq. 31 describes the role of stochastic fluctuations. The square of this term is the infinitesimal (probabilistic) variance of the changes in statistical variance $d\sigma_f^2$ of f. Just like in the stochastic replicator-mutator and Price equations, this term can profoundly affect the long-term behaviour (stationary distribution) of σ_f^2 through the slow mechanism of noise-inducec selection. Just like the stochastic Price equation, the SDE in Eq. 31 is always one-dimensional, and thus the stationary distribution of the variance σ_f^2 can also always be studied the way we studied Eq. 14.

In the case of one-dimensional quantitative traits, an infinite-dimensional version of Eq. 31 has recently been rigorously derived (Week et al., 2021) using measure-theoretic tools under certain additional assumptions (See equation (21c) in Week et al., 2021). Taking expectations over the probability space in Eq. 31 also recovers an equation previously derived and used (Débarre and Otto, 2016) in the context of evolutionary branching in finite populations as a special case (Equation A.23 in Débarre and Otto, 2016 is equivalent to our Eq. 31 for their choice of functional forms upon converting their change in variance to an infinitesimal rate of change *i.e.* derivative). An infinite population ($K \to \infty$) version of Eq. 31 also appears in Lion, 2018 (see section S7 in the online supplementary) as a dynamic version of earlier, dynamically insufficient equations for the change in trait variation over a single generation (For example, see Eq. 6.14 in Rice, 2004).

Loss of trait variation in populations experiencing genetic drift

The $\overline{\tau}\sigma_f^2$ term quantifies the loss of variation due to stochastic extinctions (i.e. demographic stochasticity) and thus represents the classic effect of neutral genetic (or ecological) drift in finite populations. To see this, it is instructive to consider the case in which this is the only force at play. Let us imagine a population of asexual organisms in which each f_i is simply a label or mark arbitrarily assigned to individuals in the population at the start of an experiment/observational study and subsequently passed on to offspring — for example, a neutral genetic tag in a part of the genome that experiences a negligible mutation rate. Since the labels are arbitrary and have no effect whatsoever on the biology of the organisms, each label has the same fitness $w_i \equiv w$ and per-capita turnover $\tau_i \equiv \tau$, and thus $\overline{w} = w$ and $\overline{\tau} = \tau$. Note that since every type has the same fitness and turnover rate, we have $\text{Cov}\left(w, (f - \overline{f})^2\right) \equiv \text{Cov}\left(\tau, (f - \overline{f})^2\right) \equiv 0$. Since the labels do not change over time, we also have $\text{Cov}\left(\partial f / \partial t, f\right) = 0$. From Eq. 31, we see that in this case, the variance changes as

$$d\sigma_f^2 = -\frac{\tau\sigma_f^2}{KN_K(t)}dt + \frac{1}{\sqrt{KN_K(t)}}dW_{\sigma_f^2} \tag{33}$$

Taking expectations, the second term on the RHS vanishes, and we see that the expected variance in the population obeys

$$\frac{d\mathbb{E}[\sigma_f^2]}{dt} = -\left(\mathbb{E}\left[\frac{\tau}{KN_K}\right]\right)\mathbb{E}[\sigma_f^2] \tag{34}$$

where we have decomposed the expectation of the product on the RHS into a product of expectations, which is admissible since the label f is completely arbitrary and thus independent of both $\overline{\tau}$ and $N_K(t)$. Eq. 34 is a simple linear ODE and has the solution

$$\mathbb{E}[\sigma_f^2](t) = \sigma_f^2(0)e^{-\mathbb{E}\left[\frac{\tau}{KN_K}\right]t} \tag{35}$$

This equation tells us that the expected diversity (variance) of labels in the population decreases exponentially over time. The rate of loss is $\mathbb{E}\left[\tau(KN_K)^{-1}\right]$, and thus, populations with higher turnover rate τ and/or lower population size KN_K lose diversity faster. This is because populations with higher τ experience more stochastic events per unit time and are thus more prone to stochastic extinction, while extinction is 'easier' in smaller populations because a smaller number of deaths is sufficient to eliminate a label from the population completely. Note that *which* labels/individuals are lost is entirely random (since all labels are arbitrary), but nevertheless, labels tend to be stochastically lost until only a single label remains in the population. Upon rescaling time as $t \to \tau t$, equation 35 recovers the continuous time version of the loss of heterozygosity formula for finite populations from population genetics (Ewens, 2004, Eq. 1.5; Crow and Kimura, 1970, sections 7.3 and 8.4).

Discussion

The central result of our paper is a set of stochastic dynamical equations for changes in trait frequencies in the population (Eq. 10) that generalizes the replicator-mutator equation to finite populations of non-constant size evolving in continuous time. From this, we derive a generalization of the Price equation (Eq. 26) and Fisher's fundamental theorem (Eq. 28) to such populations, as well as an equation for changes in population variance of a type-level quantity (Eq. 31). Starting from an arbitrary density-dependent birth-death process, our framework recovers, from first principles, well-known equations of population biology, such as the replicator-mutator equation, Price equation, and Fisher's fundamental theorem, in the infinite population limit (see section S7). Our generalised equations also reveal a novel directional evolutionary force termed noise-induced selection that emerges, surprisingly, purely from stochastic effects. Noise-induced selection can manifest through two distinct mechanisms (Box 2), one that is visible over both ecological and evolutionary timescales and one that is only visible over very long (evolutionary) timescales. Several theorists have called for a reformulation of eco-evolutionary dynamics starting from stochastic birth-death processes on the grounds that such a formulation is more fundamental and mechanistic (Metcalf and Pavard, 2007; Lambert, 2010; Doebeli et al., 2017). Our equations provide a starting point for such a reformulation by deriving some fundamental equations for the eco-evolutionary dynamics of finite, stochastically fluctuating populations. Below, we discuss some potentially fundamental implications of these equations to our understanding of stochastic evolutionary dynamics in finite populations.

Finite population effects on eco-evolutionary dynamics

For finite populations, our stochastic eco-evolutionary equations generically predict a novel directional evolutionary force called noise-induced selection that has been previously reported in model-specific contexts (Constable et al., 2016; McLeod and Day, 2019a; Parsons et al., 2018). This evolutionary force is a consequence of demographic stochasticity and can act through two distinct mechanisms (see Box 2): Over short timescales, noise-induced selection manifests as a reduction of variance in changes in population density dx_i and acts on variation in per-capita turnover rate τ, obtained by the sum of birth and death rates, of individuals. This 'fast' form of noise-induced selection is identifiable with the Gillespie effect (Gillespie, 1974) from bet-hedging theory and is detectable as a systematic deviation of the evolutionary trajectory of the system in trait frequency space relative to infinite population expectations. The fast mechanism of noise-induced selection can occur whenever there is a *differential turnover rate* τ in the system. Further, it always favors types whose per-capita turnover rate is lower than that of the population average (Box 1). Noise-induced selection can also act in a much slower fashion, visible only over very long timescales, if the variance in the change in trait frequency is itself frequency dependent. This 'slow' mechanism is only visible as a bias in the expected distribution of types over very long, evolutionary timescales (the stationary distribution), and favors types whose presence causes a reduction in the variance of changes in trait frequency dp_i. The fast and slow mechanisms may either supplement or oppose each other (see the example in the section "An example with two competing types"). Both forms of noise-induced selection can qualitatively affect the long-term evolutionary trajectory of populations and can even reverse the direction of evolution as determined by natural selection.

These results suggest an intriguing requirement for neutral evolution in finite populations: It is not sufficient for the trait in question to be neutral with respect to fitness w alone. Instead, we also require the trait to be neutral with respect to noise-induced selection. In other words, even in (finite) populations with no differential fitness among traits, there exists a directional evolutionary force that may systematically bias the course of evolution. Since noise-induced selection can act through two distinct mechanisms (see Box 2 and the example in the section "An example with two competing types"), we will see truly neutral evolution (in the sense of fixation probability only depending on initial frequency) in finite populations only when three conditions are met: no differential fitness (natural selection does not operate), no differential turnover rates (noise-induced selection does not operate through the fast mechanism), and no frequency dependence in variance of changes in trait frequency (noise-induced selection does not operate through the slow mechanism). Systematic deviations from neutrality in the absence of fitness differences have been observed in special cases before. Models of cell cycle dynamics find that selection favors cell types that periodically arrest their cell cycle (thus reducing per-capita turnover rate) relative to non-arresting cells even when their growth rates are equal (Wodarz et al., 2017). Similarly, ecological models find that when species with equal growth rates compete in finite, fluctuating populations, the species with a lower death rate outcompetes the one with a higher death rate despite the infinite population limit predicting coexistence (Lin et al., 2012; Oliveira and Dickman, 2017; Balasekaran et al., 2022). This latter bias towards species with lower death rates has sometimes been interpreted as a selection for 'longevity' (Lin et al., 2012) or a 'slower biological clock' (Oliveira and Dickman, 2017). Our analysis highlights that such results may equivalently be due to noise-induced selection preferentially favoring lower turnover rates. Our general SDEs also recover some previous work from social evolution (McLeod and Day, 2019a) and epidemiology (Parsons et al., 2018; Day et al., 2020) as special cases (Supplementary section S8).

While most evolutionary studies focus on mean traits and frequencies, we have also derived an equation for the dynamics of trait variance (Eq. 31) in finite populations. This shows that noise-induced selection can also affect the evolutionary dynamics by affecting higher moments of the trait distribution. An intriguing application of our variance equation comes from the study of evolutionary branching via adaptive dynamics. Evolutionary branching points are attracting fixed points in infinite population frameworks like adaptive dynamics (Doebeli, 2011). Since evolutionary branching is accompanied by a sudden increase in the variance of the trait in the population, it can be studied by looking at the dynamics of trait variance over time (Nordbotten et al., 2020); For finite populations, we predict that branching points of adaptive dynamics are not attractors due to the second and third terms on the RHS of Eq. 31. This observation could generically help explain why finite populations frequently exhibit a lower tendency or take a longer time to undergo evolutionary branching compared to infinite population models (Johansson and Ripa, 2006; Claessen et al., 2007; Wakano and Iwasa, 2013; Rogers and McKane, 2015; Débarre and Otto, 2016). Indeed, a special case of Eq. 31 has been used to show that evolutionary branching is less likely in finite populations in a social evolution model (Débarre and Otto, 2016).

On the practical side, the existence of noise-induced selection implies that simulation studies working with evolutionary individual-based or agent-based models should be careful about whether interaction effects are incorporated into birth rates or death rates since this seemingly arbitrary choice can have unintended consequences due to noise-induced selection, thus potentially biasing results (McLeod and Day, 2019a; Kuosmanen et al., 2022). Our results also indicate that measuring the growth rate of populations is not, in general, sufficient for accurate prediction/inference of future trajectories of the relative abundance of a species (or phenotype, allele, etc.) from empirical data even in completely controlled environments. Lastly, noise-induced selection is particular to fluctuating populations and does not occur in models with fixed population sizes such as the Wright-Fisher or Moran models. Taken alongside other theoretical (Lambert, 2010; Parsons et al., 2010; Abu Awad and Coron, 2018; Kuosmanen et al., 2022; Mazzolini and Grilli, 2023) and empirical (Papkou et al., 2016; Chavhan et al., 2019) studies on evolution in fluctuating populations, this last point suggests that models which assume fixed total population size, such as Wright-Fisher and Moran, may miss out on important evolutionary phenomena that are only seen in finite, populations of *non-constant size*.

Connections with other general frameworks

Our equations generalize Lion's (2018) general framework of infinite population deterministic eco-evolutionary dynamics to finite, fluctuating populations — taking $K \to \infty$ in Eq. 10, Eq. 26, and Eq. 31 recover equations (6), (11), and (14) in Lion, 2018 respectively. Lion, 2018 has pointed out that in the dynamic setting (for infinite populations), the replicator-mutator equation is the truly fundamental equation, and equations like the Price equation are best viewed as an infinite hierarchy of moment equations for the population mean, population variance, and higher moments of a type-level quantity. This is also true in our framework - Eq. 10 is the fundamental equation for population dynamics, and equations like Eq. 26 and Eq. 31 can then be derived from Eq. 10 through repeated application of Itô's formula. If we assume that the quantity f follows a Gaussian distribution, then the mean and variance completely characterize the distribution, and thus, Eq. 10, Eq. 26, and Eq. 31 together specify the complete stochastic dynamics of the system.

Rice has proposed a stochastic version of the Price equation (Rice, 2020 and references cited therein). Like the original Price equation, Rice's equations are formulated as a general decomposition of the phenotypic change between two given populations. They are thus the true stochastic analog of the original Price equation, whereas our version, Eq. 26, is the analog of Lion's (2018) version of the Price equation in a continuous time, dynamically sufficient setting. Rice's derivations also treat fitness as fundamental, whereas we derive suitable notions of fitness and turnover from demographic first principles. As a consequence, the 'extra' stochastic term corresponding to noise-induced selection that appears in our equations fundamentally emerges from the stochasticity of the underlying births and deaths of organisms and is thus of ecological/demographic origin, whereas the 'extra' stochastic term in Rice's equations emerges from the stochasticity of fitness alone when viewed as a random variable (Rice, 2020). It thus need not, to the best of our knowledge, correspond to the same effect we identify here.

At first glance, the idea of an evolutionary force that selects individuals with lower birth and death rates over individuals with higher birth and death rates may be reminiscent of ideas from life-history evolution such as r vs. K selection (Pianka, 1972; Stearns, 1977), bet-hedging & dormancy in temporally variable environments (Frank and Slatkin, 1990; Childs et al., 2010; Gremer and Venable, 2014; ten Brink et al., 2020), or 'pace-of-life syndromes' (Mathot and Frankenhuis, 2018; Wright et al., 2019). However, it is unclear whether these similarities reflect a deep biological principle or whether the semblance is just superficial. For one, incorporating a highly variable (non-Poisson) clutch size requires the use of stochastic processes that cannot be expressed in terms of the birth-death processes we study in this paper (for example, see Wang et al., 2023). Secondly, many models in life-history theory, such as in r vs K selection models or models of the evolution of dormancy, are often primarily concerned with spatio-temporally fluctuating external environments, and in such situations, the stochasticity in those models is extrinsic to the population (Stearns, 1977; Childs et al., 2010). We have entirely neglected such extrinsic factors in our formalism. If the variation of the environment has some associated stochasticity, the complete dynamics of the system would be the result of interactions between two qualitatively different forms of noise — *extrinsic* noise from the environment, and *intrinsic* noise from the finiteness of the population — and can consequently be rather complex and intricate (Gokhale and Hauert, 2016; Chavhan et al., 2021). Thus, while connecting noise-induced selection with ideas such as the pace-of-life syndrome is biologically appealing, it is likely a non-trivial extension and may present a promising avenue for future work.

Our equations reveal how noise-induced effects on the eco-evolutionary dynamics of populations of non-constant size can be differentiated into two major, qualitatively different effects (Box 2): The dt terms of our SDEs all contain a 'noise-induced' term that influences the expected trajectory over infinitesimal time intervals dt (The second term on the RHS of Eqs. 10, 26, and 31). This term models the expected effects of 'noise-induced selection' over very small time intervals dt, and has been the object of study in early models of bet-hedging in finite populations (Gillespie, 1974; Shpak, 2005). We can thus identify this 'fast mechanism' of noise-induced selection as the 'Gillespie effect' from the bet-hedging literature (Gillespie, 1974; Gillespie, 1977; see Box 1). However, note that due to potential additional influx terms λQ_i, the quantity τ_i is not precisely equal to the variance in the per-capita growth rate in our model, but is still proportional to it. Unlike many classic bet-hedging papers such as Gillespie (1974) and Frank and Slatkin (1990), w_i and τ_i (and thus the mean and variance of the change in population density) cannot vary independently in our framework; Instead, both are defined from first principles in terms of birth and death rates and are deeply related to each other, as is evident from their definitions (Eqs 3 and 4). This naturally introduces a tradeoff between increasing w and decreasing τ. Furthermore, the variance that is studied in bet-hedging models is typically variance in *offspring numbers* (Gillespie, 1977). The variance in Eq. 9b is *not* variance in offspring numbers, but instead variance in the ecological 'growth rate' dx_i (over an infinitesimal time interval), a quantity that has sometimes been called 'demographic variance' (Engen et al., 1998; Shpak, 2007).

Noise-induced selection over infinitesimal time scales via the fast mechanism may also appear in the 'deterministic' term (dt term) as a deviation from the expected trajectory if we project the ecological dynamics onto a 'slow manifold' through a separation of timescales argument, a common procedure for reducing the dimension of stochastic dynamical systems (Constable et al., 2013; Parsons and Rogers, 2017). A change of variables via a projection of the dynamics onto a manifold is responsible for the 'noise-induced effects' that appear in purely ecological models (i.e. models of population densities) where dynamics are projected onto a manifold describing populations that are at equilibrium over short timescales (Constable et al., 2016; Chotibut and Nelson, 2017; Mazzolini and Grilli, 2023). A change of variables via projection onto a manifold is also at the heart of the stochastic 'drift-induced selection' that drives evolutionary transitions between male and female heterogamety (XX/XY to ZW/ZZ and vice versa) in stochastic models of the evolution of chromosomal sex determination systems (Veller et al., 2017; Saunders et al., 2018). In models of sex determination, the projection is onto a manifold describing populations in which the sex ratio is 1:1 (Veller et al., 2017; Saunders et al., 2018). Thus, the effects of demographic stochasticity in these studies can be identified as manifestations of noise-induced selection acting via the fast mechanism (as defined in Box 2).

There is an entirely different effect of demographic stochasticity that is less evident over short timescales but is revealed as a systematic bias in the (quasi-)stationary distribution (see Box 2). This latter effect can occur even when the fast mechanism is absent (to see this, set $s = -\kappa / K N_K$ in Eq. 14 but let τ_1 and κ be frequency-dependent) and is a manifestation of frequency-dependence in the variance of the changes in trait frequency dp_i. This effect favors changes in frequency that lead to a reduction in the (infinitesimal) variance of the frequency process Eq. 10. To the best of our knowledge, potential directional biases in the stationary distribution due to noise-induced effects were first recognized by McLeod and Day (2019a) in the context of social evolution models of the evolution of altruism.

Since the slow and fast mechanisms of noise-induced selection have distinct origins, are visible over different timescales, and may push evolution in different directions, we suggest it is helpful to explicitly differentiate between the two mechanisms of noise-induced selection to identify which mechanisms are germane to any particular biological population (Box 2). By re-deriving some standard equations of population dynamics for finite populations, our paper presents a general description of how noise-induced selection through the fast mechanism of selection for reduced turnover rates τ_i and the slow mechanism for reduced variance in changes in trait frequency interact with each other as well as with the more well-understood evolutionary forces of natural selection, influx (mutation/migration), and neutral genetic/ecological drift. We thus provide a framework with which to approach particular finite population systems and systematically determine which evolutionary forces are important in specific contexts.

The equivalent of our stochastic equations has also recently been derived for quantitative traits from a very different starting point using the theory of measure-valued branching processes (Week et al., 2021) — equations (21b) and (21c) in Week et al., 2021 are exactly the $m \to \infty$ version of our equations for changes in the mean value of a type-level quantity and changes in the variance of a type-level quantity respectively for the special case in which the type-level quantity is the value of the quantitative trait being studied. A recent preprint (Kuosmanen et al., 2022) has also independently arrived at the equivalent of some of our equations using an alternative approximation scheme. Using certain discrete time stochastic processes and their approximation via techniques reminiscent of numerical stochastic integration, Kuosmanen et al., 2022 have arrived at our equation for type frequencies (Eq. 10) and the change of mean fitness and turnover in the population (Eq. 28 and Eq. 30), and interpreted these equations in the context of life-history evolution. However, unlike in our work, neither of these previous studies explicitly delineates or studies the interplay between the two distinct mechanisms (Box 2) through which noise-induced selection alters evolutionary dynamics.

Concluding remarks

A small but growing body of literature has begun to highlight the surprising and counter-intuitive effects of demographic stochasticity in shaping evolutionary outcomes in many ecological scenarios. In this paper, we derive from demographic first principles some general stochastic dynamical equations that conceptually unify such previous studies by connecting their dynamics to standard equations of population biology such as the replicator equation. Further, the terms of the equations we derive lend themselves to simple biological interpretations, recover standard equations of evolutionary theory in the infinite population limit, and illustrate two distinct mechanisms through which demographic stochasticity can affect evolutionary trajectories in a biased manner. The equations we derive thus provide some general insights into how evolution should operate in finite, fluctuating populations. Our results show that alongside natural selection and neutral genetic drift, finite populations also experience an additional *directional* force (noise-induced selection) that can affect evolutionary trajectories, sometimes even reversing the direction of evolution predicted by natural selection. Furthermore, this noise-induced selection can operate through either a 'fast' mechanism that is visible over short timescales or a 'slow' mechanism that is visible only over very long timescales as a biasing of the stationary distribution. To the best of our knowledge, the equations we derive in this paper are the first to showcase how noise-induced selection alters some standard equations of population biology. The utility of the equations we derive thus lies not (necessarily) in their solutions for specific models, but instead in their generality and the fact that their terms help us clearly think about the various evolutionary phenomena operating in biological populations (Queller, 2017; Lehtonen, 2018; Lion, 2018; Luque and Baravalle, 2021). We work solely with unstructured populations in this paper, thus neglecting any potential effects of groups, sex, age, or space. Future work could potentially focus on including these additional features to investigate their effects in amplifying or reducing the effects of noise-induced selection.

Acknowledgements

We are grateful to Srikanth Iyer, Kavita Jain, Sébastien Lion, and two anonymous reviewers for their helpful comments on an earlier draft of this manuscript. ASB is supported by a Kishore Yaigyanik Protsahan Yojana (KVPY) fellowship from the Department of Science and Technology of the Government of India (Fellowship ID: SX-1711025). VG is supported by grants from the Science and Engineering Research Board (SERB), Government of India (MTR/2022/000273).

Author Contributions

Ananda Shikhara Bhat: Conceptualization, Methodology, Formal Analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization; **Vishwesha Guttal:** Conceptualization, Methodology, Validation, Writing - Review & Editing, Supervision.

Literature Cited

Abu Awad, D. and Coron, C. (2018). “Effects of demographic stochasticity and life-history strategies on times and probabilities to fixation”. en. In: *Heredity* 121.4, pp. 374–386. ISSN: 1365-2540. DOI: `10.1038/s41437-018-0118-6`.

Alexander, H. and Wahl, L. (2008). “Fixation Probabilities Depend on Life History: Fecundity, Generation time and Survival in a burst-death model”. In: *Evolution* 62.7, pp. 1600–1609. ISSN: 0014-3820. DOI: `10.1111/j.1558-5646.2008.00396.x`.

Balasekaran, M., Johanis, M., Rychtář, J., Taylor, D., and Zhu, J. (2022). “Quasi-neutral evolution in populations under small demographic fluctuations”. en. In: *Journal of Theoretical Biology* 538, p. 111040. ISSN: 0022-5193. DOI: `10.1016/j.jtbi.2022.111040`.

Black, A. J. and McKane, A. J. (2012). “Stochastic formulation of ecological models and their applications”. en. In: *Trends in Ecology & Evolution* 27.6, pp. 337–345. ISSN: 0169-5347. DOI: `10.1016/j.tree.2012.01.014`.

Boettiger, C. (2018). “From noise to knowledge: how randomness generates novel phenomena and reveals information”. en. In: *Ecology Letters* 21.8, pp. 1255–1267. ISSN: 1461-0248. DOI: `10.1111/ele.13085`.

Ten Brink, H., Gremer, J. R., and Kokko, H. (2020). “Optimal germination timing in unpredictable environments: the importance of dormancy for both among- and within-season variation”. en. In: *Ecology Letters* 23.4, pp. 620–630. ISSN: 1461-0248. DOI: `10.1111/ele.13461`.

Chavhan, Y., Ali, S. I., and Dey, S. (2019). “Larger Numbers Can Impede Adaptation in Asexual Populations despite Entailing Greater Genetic Variation”. en. In: *Evolutionary Biology* 46.1, pp. 1–13. ISSN: 1934-2845. DOI: `10.1007/s11692-018-9467-6`.

Chavhan, Y., Malusare, S., and Dey, S. (2021). “Interplay of population size and environmental fluctuations: A new explanation for fitness cost rarity in asexuals”. en. In: *Ecology Letters* 24.9, pp. 1943–1954. ISSN: 1461-0248. DOI: `10.1111/ele.13831`.

Childs, D. Z., Metcalf, C. J. E., and Rees, M. (2010). “Evolutionary bet-hedging in the real world: empirical evidence and challenges revealed by plants”.

In: *Proceedings of the Royal Society B: Biological Sciences* 277.1697, pp. 3055–3064. DOI: 10.1098/rspb.2010.0707.

Chotibut, T. and Nelson, D. R. (2015). "Evolutionary dynamics with fluctuating population sizes and strong mutualism". In: *Physical Review E* 92.2, p. 022718. DOI: 10.1103/PhysRevE.92.022718.

— (2017). "Population Genetics with Fluctuating Population Sizes". en. In: *Journal of Statistical Physics* 167.3, pp. 777–791. ISSN: 1572-9613. DOI: 10.1007/s10955-017-1741-y.

Claessen, D., Andersson, J., Persson, L., and Roos, A. M. d. (2007). "Delayed evolutionary branching in small populations". English. In: *Evolutionary Ecology Research* 9.1, pp. 51–69. ISSN: 1522-0613.

Constable, G. W. A., McKane, A. J., and Rogers, T. (2013). "Stochastic dynamics on slow manifolds". en. In: *Journal of Physics A: Mathematical and Theoretical* 46.29, p. 295002. ISSN: 1751-8121. DOI: 10.1088/1751-8113/46/29/295002.

Constable, G. W. A., Rogers, T., McKane, A. J., and Tarnita, C. E. (2016). "Demographic noise can reverse the direction of deterministic selection". In: *Proceedings of the National Academy of Sciences* 113.32, E4745–E4754. DOI: 10.1073/pnas.1603693113.

Coulson, T., Rohani, P., and Pascual, M. (2004). "Skeletons, noise and population growth: the end of an old debate?" en. In: *Trends in Ecology & Evolution* 19.7, pp. 359–364. ISSN: 0169-5347. DOI: 10.1016/j.tree.2004.05.008.

Crow, J. F. and Kimura, M. (1970). *An introduction to population genetics theory*. eng. Harper international editions. New York: Harper & Row.

Czuppon, P. and Traulsen, A. (2021). "Understanding evolutionary and ecological dynamics using a continuum limit". en. In: *Ecology and Evolution* 11.11, pp. 5857–5873. ISSN: 2045-7758. DOI: 10.1002/ece3.7205.

Day, T., Parsons, T. L., Lambert, A., and Gandon, S. (2020). "The Price equation and evolutionary epidemiology". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1797, p. 20190357. DOI: 10.1098/rstb.2019.0357.

Débarre, F. and Otto, S. P. (2016). "Evolutionary dynamics of a quantitative trait in a finite asexual population". en. In: *Theoretical Population Biology* 108, pp. 75–88. ISSN: 0040-5809. DOI: 10.1016/j.tpb.2015.12.002.

DeLong, J. P. and Cressler, C. E. (2023). “Stochasticity directs adaptive evolution toward nonequilibrium evolutionary attractors”. en. In: *Ecology* 104.1, e3873. ISSN: 1939-9170. DOI: 10.1002/ecy.3873.

Doebeli, M. (2011). *Adaptive diversification*. English. Princeton, N.J.: Princeton University Press. ISBN: 978-0-691-12894-8.

Doebeli, M., Ispolatov, Y., and Simon, B. (2017). “Towards a mechanistic foundation of evolutionary theory”. In: *eLife* 6. Ed. by Wenying Shou, e23804. ISSN: 2050-084X. DOI: 10.7554/eLife.23804.

Engen, S., Bakke, Ø., and Islam, A. (1998). “Demographic and Environmental Stochasticity-Concepts and Definitions”. In: *Biometrics* 54.3, pp. 840–846. ISSN: 0006-341X. DOI: 10.2307/2533838.

Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes: characterization and convergence*. eng. Wiley series in probability and mathematical statistics. New York: Wiley. ISBN: 978-0-471-08186-9.

Ewens, W. J. (2004). *Mathematical Population Genetics 1: Theoretical Introduction*. en. Springer Science & Business Media. ISBN: 978-0-387-20191-7.

Frank, S. A. (2012). “Natural selection. IV. The Price equation*”. en. In: *Journal of Evolutionary Biology* 25.6, pp. 1002–1019. ISSN: 1420-9101. DOI: 10.1111/j.1420-9101.2012.02498.x.

Frank, S. A. and Slatkin, M. (1990). “Evolution in a Variable Environment”. In: *The American Naturalist* 136.2, pp. 244–260. ISSN: 0003-0147. DOI: 10.1086/285094.

— (1992). “Fisher’s fundamental theorem of natural selection”. en. In: *Trends in Ecology & Evolution* 7.3, pp. 92–95. ISSN: 0169-5347. DOI: 10.1016/0169-5347(92)90248-A.

Fronhofer, E. A., Corenblit, D., Deshpande, J. N., Govaert, L., Huneman, P., Viard, F., Jarne, P., and Puijalon, S. (2023). “Eco-evolution from deep time to contemporary dynamics: The role of timescales and rate modulators”. en. In: *Ecology Letters* 26.S1, S91–S108. ISSN: 1461-0248. DOI: 10.1111/ele.14222.

Gillespie, J. H. (1974). “Natural selection for within-generation variance in offspring number”. In: *Genetics* 76.3, pp. 601–606. ISSN: 1943-2631. DOI: 10.1093/genetics/76.3.601.

— (1977). “Natural Selection for Variances in Offspring Numbers: A New Evolutionary Principle”. In: *The American Naturalist* 111.981, pp. 1010–1014. ISSN: 0003-0147. DOI: 10.1086/283230.

Gokhale, C. S. and Hauert, C. (2016). “Eco-evolutionary dynamics of social dilemmas”. en. In: *Theoretical Population Biology* 111, pp. 28–42. ISSN: 0040-5809. DOI: 10.1016/j.tpb.2016.05.005.

Gremer, J. R. and Venable, D. L. (2014). “Bet hedging in desert winter annual plants: optimal germination strategies in a variable environment”. en. In: *Ecology Letters* 17.3, pp. 380–387. ISSN: 1461-0248. DOI: 10.1111/ele.12241.

Horsthemke, W. and Lefever, R. (1984). *Noise-induced transitions: theory and applications in physics, chemistry and biology*. Springer series in synergetics. Berlin: Springer-Verlag. ISBN: 978-3-540-11359-1.

Houchmandzadeh, B. and Vallade, M. (2012). “Selection for altruism through random drift in variable size populations”. In: *BMC Evolutionary Biology* 12.1, p. 61. ISSN: 1471-2148. DOI: 10.1186/1471-2148-12-61.

Humplik, J., Hill, A. L., and Nowak, M. A. (2014). “Evolutionary dynamics of infectious diseases in finite populations”. en. In: *Journal of Theoretical Biology* 360, pp. 149–162. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2014.06.039.

Jhawar, J., Morris, R. G., Amith-Kumar, U. R., Danny Raj, M., Rogers, T., Rajendran, H., and Guttal, V. (2020). “Noise-induced schooling of fish”. en. In: *Nature Physics* 16.4, pp. 488–493. ISSN: 1745-2481. DOI: 10.1038/s41567-020-0787-y.

Johansson, J. and Ripa, J. (2006). “Will Sympatric Speciation Fail due to Stochastic Competitive Exclusion?” In: *The American Naturalist* 168.4, pp. 572–578. ISSN: 0003-0147. DOI: 10.1086/507996.

Joshi, J. and Guttal, V. (2018). “Demographic noise and cost of greenbeard can facilitate greenbeard cooperation”. en. In: *Evolution* 72.12, pp. 2595–2607. ISSN: 1558-5646. DOI: 10.1111/evo.13615.

Karlin, S. and Taylor, H. M. (1981). *A second course in stochastic processes*. eng. New York: Academic Press. ISBN: 978-0-12-398650-4.

Kimura, M. and Ohta, T. (1974). “Probability of Gene Fixation in an Expanding Finite Population”. In: *Proceedings of the National Academy of Sciences* 71.9, pp. 3377–3379. DOI: 10.1073/pnas.71.9.3377.

Kokko, H. (2021). “The stagnation paradox: the ever-improving but (more or less) stationary population fitness”. In: *Proceedings of the Royal Society B: Biological Sciences* 288.1963, p. 20212145. DOI: 10.1098/rspb.2021.2145.

Kuosmanen, T., Särkkä, S., and Mustonen, V. (2022). *Turnover shapes evolution of birth and death rates*. en. DOI: 10.1101/2022.07.11.499527.

Lambert, A. (2010). “Population genetics, ecology and the size of populations”. en. In: *Journal of Mathematical Biology* 60.3, pp. 469–472. ISSN: 1432-1416. DOI: 10.1007/s00285-009-0286-3.

Lande, R. (1976). "Natural Selection and Random Genetic Drift in Phenotypic Evolution". In: *Evolution* 30.2, pp. 314–334. ISSN: 0014-3820. DOI: 10.2307/2407703.

Lehtonen, J. (2018). "The Price Equation, Gradient Dynamics, and Continuous Trait Game Theory". In: *The American Naturalist* 191.1, pp. 146–153. ISSN: 0003-0147. DOI: 10.1086/694891.

— (2020). "The Price equation and the unity of social evolution theory". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1797, p. 20190362. DOI: 10.1098/rstb.2019.0362.

Lin, Y. T., Kim, H., and Doering, C. R. (2012). "Features of Fast Living: On the Weak Selection for Longevity in Degenerate Birth-Death Processes". en. In: *Journal of Statistical Physics* 148.4, pp. 647–663. ISSN: 1572-9613. DOI: 10.1007/s10955-012-0479-9.

Lion, S. (2018). "Theoretical Approaches in Evolutionary Ecology: Environmental Feedback as a Unifying Perspective". In: *The American Naturalist* 191.1, pp. 21–44. ISSN: 0003-0147. DOI: 10.1086/694865.

Luque, V. J. and Baravalle, L. (2021). "The mirror of physics: on how the Price equation can unify evolutionary biology". en. In: *Synthese* 199.5, pp. 12439–12462. ISSN: 1573-0964. DOI: 10.1007/s11229-021-03339-6.

Majumder, S., Das, A., Kushal, A., Sankaran, S., and Guttal, V. (2021). "Finite-size effects, demographic noise, and ecosystem dynamics". en. In: *The European Physical Journal Special Topics* 230.16, pp. 3389–3401. ISSN: 1951-6401. DOI: 10.1140/epjs/s11734-021-00184-z.

Mathot, K. J. and Frankenhuis, W. E. (2018). "Models of pace-of-life syndromes (POLS): a systematic review". en. In: *Behavioral Ecology and Sociobiology* 72.3, p. 41. ISSN: 1432-0762. DOI: 10.1007/s00265-018-2459-9.

Mazzolini, A. and Grilli, J. (2023). "Universality of evolutionary trajectories under arbitrary forms of self-limitation and competition". In: *Physical Review E* 108.3, p. 034406. DOI: 10.1103/PhysRevE.108.034406.

McLeod, D. V. and Day, T. (2019a). "Social evolution under demographic stochasticity". en. In: *PLOS Computational Biology* 15.2, e1006739. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006739.

— (2019b). "Why is sterility virulence most common in sexually transmitted infections? Examining the role of epidemiology". In: *Evolution* 73.5, pp. 872–882. ISSN: 0014-3820. DOI: 10.1111/evo.13718.

Metcalf, C. J. E. and Pavard, S. (2007). "Why evolutionary biologists should be demographers". en.

In: *Trends in Ecology & Evolution* 22.4, pp. 205–212. ISSN: 0169-5347. DOI: 10.1016/j.tree.2006.12.001.

Nordbotten, J. M., Bokma, F., Hermansen, J. S., and Stenseth, N. C. (2020). “The dynamics of trait variance in multi-species communities”. In: *Royal Society Open Science* 7.8, p. 200321. DOI: 10.1098/rsos.200321.

Oliveira, M. M. d. and Dickman, R. (2017). “The advantage of being slow: The quasi-neutral contact process”. en. In: *PLOS ONE* 12.8, e0182672. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0182672.

Olofsson, H., Ripa, J., and Jonzén, N. (2009). “Bet-hedging as an evolutionary game: the trade-off between egg size and number”. In: *Proceedings of the Royal Society B: Biological Sciences* 276.1669, pp. 2963–2969. DOI: 10.1098/rspb.2009.0500.

Page, K. M. and Nowak, M. A. (2002). “Unifying Evolutionary Dynamics”. en. In: *Journal of Theoretical Biology* 219.1, pp. 93–98. ISSN: 0022-5193. DOI: 10.1006/jtbi.2002.3112.

Papkou, A., Gokhale, C. S., Traulsen, A., and Schulenburg, H. (2016). “Host–parasite coevolution: why changing population size matters”. In: *Zoology*. SI: Host-Parasite Coevolution 119.4, pp. 330–338. ISSN: 0944-2006. DOI: 10.1016/j.zool.2016.02.001.

Parsons, T. L., Lambert, A., Day, T., and Gandon, S. (2018). “Pathogen evolution in finite populations: slow and steady spreads the best”. In: *Journal of The Royal Society Interface* 15.147, p. 20180135. DOI: 10.1098/rsif.2018.0135.

Parsons, T. L., Quince, C., and Plotkin, J. B. (2010). “Some Consequences of Demographic Stochasticity in Population Genetics”. In: *Genetics* 185.4, pp. 1345–1354. ISSN: 1943-2631. DOI: 10.1534/genetics.110.115030.

Parsons, T. L. and Rogers, T. (2017). “Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: a constructive approach with examples from theoretical biology”. en. In: *Journal of Physics A: Mathematical and Theoretical* 50.41, p. 415601. ISSN: 1751-8121. DOI: 10.1088/1751-8121/aa86c7.

Pianka, E. R. (1972). “r and K Selection or b and d Selection?” In: *The American Naturalist* 106.951, pp. 581–588. ISSN: 0003-0147. DOI: 10.1086/282798.

Proulx, S. R. and Day, T. (2005). “What can Invasion Analyses Tell us about Evolution under Stochasticity in Finite Populations?” In: *Selection* 2.1-2, pp. 2–15. ISSN: 1585-1931, 1588-287X. DOI: 10.1556/select.2.2001.1-2.2.

Queller, D. C. (2017). “Fundamental Theorems of Evolution”. In: *The American Naturalist* 189.4, pp. 345–353. ISSN: 0003-0147. DOI: 10.1086/690937.

Raatz, M. and Traulsen, A. (2023). *Promoting extinction or minimizing growth? The impact of treatment on trait trajectories in evolving populations*. en. DOI: `10.1101/2022.06.17.496570`.

Rice, S. H. (2004). *Evolutionary theory: mathematical and conceptual foundations*. English. Sunderland, Mass., USA: Sinauer Associates. ISBN: 978-0-87893-702-8.

— (2020). "Universal rules for the interaction of selection and transmission in evolution". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1797, p. 20190353. DOI: `10.1098/rstb.2019.0353`.

Rogers, T. and McKane, A. J. (2015). "Modes of competition and the fitness of evolved populations". In: *Physical Review E* 92.3, p. 032708. DOI: `10.1103/PhysRevE.92.032708`.

Saunders, P. A., Neuenschwander, S., and Perrin, N. (2018). "Sex chromosome turnovers and genetic drift: a simulation study". en. In: *Journal of Evolutionary Biology* 31.9, pp. 1413–1419. ISSN: 1420-9101. DOI: `10.1111/jeb.13336`.

Seger, J. and Brockmann, H. J. (1987). "What is bet-hedging?" In: *Oxford surveys in evolutionary biology*. Ed. by P.H. Harvey and L Partridge. Vol. 4. Oxford University Press, pp. 182–211.

Shpak, M. (2005). "Evolution of variance in offspring number: The effects of population size and migration". en. In: *Theory in Biosciences* 124.1, pp. 65–85. ISSN: 1611-7530. DOI: `10.1016/j.thbio.2005.05.003`.

— (2007). "Selection Against Demographic Stochasticity in Age-Structured Populations". In: *Genetics* 177.4, pp. 2181–2194. ISSN: 1943-2631. DOI: `10.1534/genetics.107.080747`.

Starrfelt, J. and Kokko, H. (2012). "Bet-hedging—a triple trade-off between means, variances and correlations". en. In: *Biological Reviews* 87.3, pp. 742–755. ISSN: 1469-185X. DOI: `10.1111/j.1469-185X.2012.00225.x`.

Stearns, S. C. (1977). "The Evolution of Life History Traits: A Critique of the Theory and a Review of the Data". In: *Annual Review of Ecology and Systematics* 8.1, pp. 145–171. DOI: `10.1146/annurev.es.08.110177.001045`.

Strang, A. G., Abbott, K. C., and Thomas, P. J. (2019). "How to avoid an extinction time paradox". en. In: *Theoretical Ecology* 12.4, pp. 467–487. ISSN: 1874-1746. DOI: `10.1007/s12080-019-0416-5`.

Van Kampen, N. G. (1981). *Stochastic processes in physics and chemistry*. eng. Amsterdam, New York, New York: North-Holland. ISBN: 978-0-444-86200-6.

Veller, C., Muralidhar, P., Constable, G. W. A., and Nowak, M. A. (2017). "Drift-Induced Selection Between Male and Female Heterogamety". en. In: *Genetics* 207.2, pp. 711–727. ISSN: 1943-2631. DOI: `10.1534/genetics.117.300151`.

Wakano, J. Y. and Iwasa, Y. (2013). "Evolutionary Branching in a Finite Population: Deterministic Branching vs. Stochastic Branching". In: *Genetics* 193.1, pp. 229–241. ISSN: 1943-2631. DOI: `10.1534/genetics.112.144980`.

Wang, G., Su, Q., Wang, L., and Plotkin, J. B. (2023). "Reproductive variance can drive behavioral dynamics". In: *Proceedings of the National Academy of Sciences* 120.12, e2216218120. DOI: `10.1073/pnas.2216218120`.

Week, B., Nuismer, S. L., Harmon, L. J., and Krone, S. M. (2021). "A white noise approach to evolutionary ecology". In: *Journal of Theoretical Biology* 521, p. 110660. ISSN: 0022-5193. DOI: `10.1016/j.jtbi.2021.110660`.

Wodarz, D., Goel, A., and Komarova, N. L. (2017). "Effect of cell cycle duration on somatic evolutionary dynamics". en. In: *Evolutionary Applications* 10.10, pp. 1121–1129. ISSN: 1752-4571. DOI: `10.1111/eva.12518`.

Wright, J., Bolstad, G. H., Araya-Ajoy, Y. G., and Dingemanse, N. J. (2019). "Life-history evolution under fluctuating density-dependent selection and the adaptive alignment of pace-of-life syndromes". en. In: *Biological Reviews* 94.1, pp. 230–247. ISSN: 1469-185X. DOI: `10.1111/brv.12451`.

Yamamichi, M., Ellner, S. P., and Hairston Jr., N. G. (2023). "Beyond simple adaptation: Incorporating other evolutionary processes and concepts into eco-evolutionary dynamics". en. In: *Ecology Letters* 26.S1, S16–S21. ISSN: 1461-0248. DOI: `10.1111/ele.14197`.

Supplementary Information for

Bhat and Guttal 2024: Eco-evolutionary dynamics for finite populations and the noise-induced reversal of selection

Ananda Shikhara Bhat[1,2,3,*], Vishwesha Guttal[1]

[1] Centre for Ecological Sciences, Indian Institute of Science, Bengaluru, Karnataka-560012, India
[2] Department of Biology, Indian Institute of Science Education and Research Pune, Maharashtra-411008, India
[3] *Current Affiliation:* Institute of Organismic and Molecular Evolution (iomE), Johannes Gutenberg University, 55128 Mainz, Germany; Institute for Quantitative and Computational Biosciences (IQCB), Johannes Gutenberg University, 55128 Mainz, Germany

[*] E-mail for correspondence: abhat@uni-mainz.de

Table of Contents

S1 The master equation and the system size expansion

Given a system with m different types of individuals and birth and death rate functions $b_i(\mathbf{n})$ and $d_i(\mathbf{n})$, we are interested in finding an equation for the rate of change of the conditional probability $P(\mathbf{n},t|\mathbf{n}_0,0)$, the probability of finding the population in a state $\mathbf{n}$ at time t. Henceforth, we omit the conditioning for notational brevity and simply write $P(\mathbf{n},t)$ for this quantity. We assume that the birth and death rates are of the order of the total population size, *i.e.* that $b_i(\mathbf{n})$ and $d_i(\mathbf{n})$ are $\mathcal{O}(\sum_i n_i)$ functions.

For each $i \in \{1,\ldots,m\}$, let us now define two step operators $\mathcal{E}_i^{\pm}$ by their action on any function $f([n_1,\ldots,n_m],t)$ as:

$$\mathcal{E}_i^{\pm} f([n_1,\ldots,n_i,\ldots,n_m],t) = f([n_1,\ldots,n_i \pm 1,\ldots n_m],t) \tag{S1}$$

In other words, $\mathcal{E}_i^{\pm}$ just changes the population through the addition or removal of one type i individual. We can now write down an exact equation for the rate of change of $P(\mathbf{n},t)$ by noting that the only direct transitions allowed are those from populations that are exactly one individual away from our focal population. Thus, we have the relation

$$\frac{\partial P}{\partial t}(\mathbf{n},t) = \sum_{j=1}^{m}\left[(\mathcal{E}_j^{-}-1)b_j(\mathbf{n})P(\mathbf{n},t) + (\mathcal{E}_j^{+}-1)d_j(\mathbf{n})P(\mathbf{n},t)\right] \tag{S2}$$

This equation is called the 'master equation', and completely characterizes our m-dimensional process.

As mentioned in the main text, we assume that there is a carrying capacity/population size measure $K > 0$ such that the total population size $\sum_i n_i$ is expected to be $\mathcal{O}(K)$. This allows us to move from population numbers $\mathbf{n}$ to population 'densities' $\mathbf{x} = \mathbf{n}/K$. Specifically, we assume that we can find $\mathcal{O}(1)$ functions $b_i^{(K)}$ and $d_i^{(K)}$ such that we can write

$$\begin{aligned} b_i(\mathbf{n}) &= K b_i^{(K)}(\mathbf{x}) \\ d_i(\mathbf{n}) &= K d_i^{(K)}(\mathbf{x}) \end{aligned} \tag{S3}$$

Note that this assumption means that $b_i^{(K)}$ and $d_i^{(K)}$ remain well-defined even in the $K \to \infty$ limit, since b_i/K and d_i/K remain $\mathcal{O}(1)$ by our assumption on the scaling properties of $\mathbf{n}$, b_i, and d_i. Thus, we may still speak of population densities x in the infinite population size limit ($K \to \infty$). Note that this scaling assumption implies that in the functional forms 2, we assume that $b_i^{(\text{ind})}(\mathbf{x})$, $d_i^{(\text{ind})}(\mathbf{x})$, and $Q_i(\mathbf{x})$ are all $\mathcal{O}(1)$ functions.

To describe our stochastic process in terms of population densities rather than absolute population sizes, we now define new step operators $\Delta_i^{\pm}$ by their action on any real-valued function

$f(\mathbf{x},t)$ as

$$\Delta_i^{\pm} f([x_1,\ldots,x_m],t) = f([x_1,\ldots,x_i \pm \frac{1}{K},\ldots x_m],t) \tag{S4}$$

In terms of these new variables, (S2) becomes

$$\frac{\partial P}{\partial t}(\mathbf{x},t) = K\sum_{j=1}^{m}\left[(\Delta_j^- - 1)b_j^{(K)}(\mathbf{x})P(\mathbf{x},t) + (\Delta_j^+ - 1)d_j^{(K)}(\mathbf{x})P(\mathbf{x},t)\right] \tag{S5}$$

If K is large, we can now Taylor expand the action of the step operators as

$$f([x_1,\ldots,x_i \pm \frac{1}{K},\ldots x_m],t) = f(\mathbf{x},t) \pm \frac{1}{K}\frac{\partial f}{\partial x_i}(\mathbf{x},t) + \frac{1}{2K^2}\frac{\partial^2 f}{\partial x_i^2}(\mathbf{x},t) + \cdots$$

which, after substituting into (S5) and neglecting higher order terms, yields the equation

$$\frac{\partial P}{\partial t}(\mathbf{x},t) = \sum_{j=1}^{m}\left[-\frac{\partial}{\partial x_j}\{A_j^-(\mathbf{x})P(\mathbf{x},t)\} + \frac{1}{2K}\frac{\partial^2}{\partial x_j^2}\{A_j^+(\mathbf{x})P(\mathbf{x},t)\}\right] \tag{S6}$$

where

$$A_i^{\pm}(\mathbf{x}) = b_i^{(K)}(\mathbf{x}) \pm d_i^{(K)}(\mathbf{x})$$

Equation (S6) is an m-dimensional version of a 'Fokker-Planck equation' for the probability density $P(x,t)$.

Itô SDE representation

For our purposes, we will often find it convenient to describe the same process as defined by the Fokker-Planck equation (S6) via an 'Itô stochastic differential equation'. It is well-known (Øksendal, 1998) that a stochastic process whose probability density function satisfies a Fokker-Planck equation of the form (S6) is equivalent to an m-dimensional stochastic process obtained as the solution to the Itô SDE

$$d\mathbf{X}_t = \mathbf{A}^-(\mathbf{X}_t)dt + \frac{1}{\sqrt{K}}\mathbf{D}(\mathbf{X}_t)d\mathbf{W}_t \tag{S7}$$

Here, $\mathbf{A}^-(\mathbf{X}_t)$ is an m-dimensional vector with i^{th} element $= A_i^-(\mathbf{X}_t)$. $\mathbf{D}(\mathbf{X}_t)$ is an $m \times m$ matrix with ijth element $(\mathbf{D}(\mathbf{X}_t))_{ij} = \delta_{ij}\left(A_i^+ A_j^+\right)^{\frac{1}{4}}$, where δ_{ij} is the Kronecker delta symbol, defined by

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Finally, $\mathbf{W}_t$ is the m-dimensional Wiener process (standard Brownian motion) and can be thought of as a vector of independent one-dimensional Wiener processes.

For a more detailed discussion on such approximations, we refer the reader to Chapter 11 of Ethier and Kurtz, 1986 for the mathematically rigorous theory and Chapter 10 of Van Kampen, 1981 for a heuristic approach. Pedagogical treatments focused on eco-evolutionary population dynamics can be found in Black and McKane, 2012 and Czuppon and Traulsen, 2021.

S2 Trait frequency dynamics using Itô's formula

We first recall the version of the multi-dimensional Itô's formula that will be relevant to us. Consider an m-dimensional real Itô process $\mathbf{X}_t$ given by the solution to

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t)dt + \boldsymbol{\sigma}(\mathbf{X}_t)d\mathbf{W}_t$$

where $\boldsymbol{\mu} : \mathbb{R}^m \to \mathbb{R}^m$ is the 'drift vector' and $\boldsymbol{\sigma} : \mathbb{R}^m \to \mathbb{R}^{m\times m}$ is the 'diffusion matrix'. Let $f : \mathbb{R}^m \to \mathbb{R}$ be an arbitrary $C^2(\mathbb{R}^m)$ function. Then, Itô's formula (Øksendal, 1998, Section 4.2) states that the stochastic process $f(\mathbf{X}_t)$ must satisfy:

$$df(\mathbf{X}_t) = \left[(\nabla_{\mathbf{X}} f)^{\mathrm{T}} \boldsymbol{\mu} + \frac{1}{2}\mathrm{Tr}[\boldsymbol{\sigma}^{\mathrm{T}}(H_{\mathbf{X}} f)\boldsymbol{\sigma}]\right] dt + (\nabla_{\mathbf{X}} f)^{\mathrm{T}} \boldsymbol{\sigma} d\mathbf{W}_t \tag{S8}$$

where $\mathrm{Tr}[\cdot]$ denotes the trace of a matrix, $(\cdot)^{\mathrm{T}}$ denotes the transpose, and we have suppressed the $\mathbf{X}_t$ dependence of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ to reduce clutter. Here, $\nabla_{\mathbf{x}} f$ is the m-dimensional *gradient vector* of f with respect to $\mathbf{x}$ and $H_{\mathbf{x}} f$ is the $m \times m$ *Hessian matrix* of f with respect to $\mathbf{x}$, respectively defined for $f([x_1, \ldots, x_m]^{\mathrm{T}})$ as:

$$(\nabla_{\mathbf{x}} f)_j = \frac{\partial f}{\partial x_j}$$

$$(H_{\mathbf{x}} f)_{jk} = \frac{\partial^2 f}{\partial x_j \partial x_k}$$

In our case, we have the Itô process given by (S7), which defines how the density of each type of individual changes over time. We thus have $\boldsymbol{\mu}(\mathbf{X}_t) = \mathbf{A}^-(\mathbf{X}_t)$ and $\boldsymbol{\sigma}(\mathbf{X}_t) = \mathbf{D}(\mathbf{X}_t)/\sqrt{K}$. For each fixed $i \in \{1, 2, \ldots, m\}$, let us define a scalar function $f_i : \mathbb{R}^m \to \mathbb{R}$ as

$$f_i(\mathbf{x}) = \frac{x_i}{\sum\limits_{j=1}^{m} x_j}$$

Thus, $f_i(\mathbf{X}_t)$ gives us the frequency of type i individuals when the population is described by the vector $\mathbf{X}_t$. This function is obviously $C^2(\mathbb{R}^m)$, and we can thus use Itô's formula (S8) to describe how it changes over time. The j^{th} element of the gradient of f_i is given by:

$$\begin{aligned}(\nabla_{\mathbf{x}} f_i)_j &= \frac{\partial}{\partial x_j}\left(\frac{x_i}{\sum\limits_{k=1}^{m} x_k}\right)\\ &= \left(\left(\frac{1}{\sum\limits_{r=1}^{m} x_r}\right)\frac{\partial x_i}{\partial x_j} - \left(\frac{x_i}{\left(\sum\limits_{r=1}^{m} x_r\right)^2}\right)\sum_{k=1}^{m}\frac{\partial x_k}{\partial x_j}\right) \qquad \text{(S9)}\end{aligned}$$

where we have defined the frequency of the i^{th} type $p_i = f_i(\mathbf{x})$. To proceed further, we require the quantity $\frac{\partial x_j}{\partial x_k}$ for any pair of types $j,k \in \{1,2,3,\ldots,m-1,m\}$. Since changes in densities in our system are only being determined by ecological interactions at the individual level, with changes in total population size being an emergent quantity, we can assume that our system obeys $\frac{\partial x_j}{\partial x_k} = \delta_{jk}\ \forall\ j,k \in \{1,2,3,\ldots,m-1,m\}$. Note that this is *not* true if the total population size is held constant since changes in densities of one type must be accompanied by complementary changes in densities of at least one other type to keep the total density $\sum_i x_i$ strictly constant.

We can now substitute $\frac{\partial x_j}{\partial x_k} = \delta_{jk}$ into equation (S9). Upon doing this, we obtain

$$(\nabla_{\mathbf{x}} f_i)_j = \frac{1}{\sum\limits_{r=1}^{m} x_r}\left(\delta_{ij} - p_i\right) \qquad \text{(S10)}$$

Similarly, we can also calculate elements of the Hessian matrix. The jk^{th} element of the Hessian is given by:

$$\begin{aligned}(H_{\mathbf{x}} f_i)_{jk} &= \frac{\partial^2}{\partial x_j \partial x_k}\left(\frac{x_i}{\sum\limits_{l=1}^{m} x_l}\right)\\ &= \frac{\partial}{\partial x_j}\left(\frac{\delta_{ik}}{\sum\limits_{r=1}^{m} x_r} - \frac{x_i}{\left(\sum\limits_{r=1}^{m} x_r\right)^2}\right)\\ &= \frac{1}{\left(\sum\limits_{r=1}^{m} x_r\right)^2}\left(2p_i - \delta_{ij} - \delta_{ik}\right) \qquad \text{(S11)}\end{aligned}$$

Thus, for the first term of (S8), we have:

$$\begin{aligned}
\left(\nabla_{\mathbf{x}} f_i\right)^{\mathrm{T}} \mathbf{A}^{-} &= \sum_{j=1}^{m} \left(\left(\nabla_{\mathbf{x}} f_i\right)_j\right) A_j^{-} \\
&= \frac{1}{\sum\limits_{r=1}^{m} x_r} \sum_{j=1}^{m} \left(\delta_{ij} - p_i\right) A_j^{-} \\
&= \frac{1}{\sum\limits_{r=1}^{m} x_r} \left(A_i^{-} - p_i \sum_{j=1}^{m} A_j^{-}\right) \qquad \text{(S12)}
\end{aligned}$$

This term describes the effects of selection and influx (mutation/migration) at the infinite population limit. However, the finiteness of the population adds a second directional term to these dynamics, described by the second term that multiplies dt in (S8). To calculate it, we first calculate:

$$\begin{aligned}
\frac{1}{\sqrt{K}} \left(H_{\mathbf{x}} f_i \mathbf{D}\right)_{jk} &= \frac{1}{\sqrt{K}} \sum_{l=1}^{m} \left(H_{\mathbf{x}} f_i\right)_{jl} \left(\mathbf{D}\right)_{lk} \\
&= \frac{1}{\sqrt{K}\left(\sum\limits_{r=1}^{m} x_r\right)^2} \sum_{l=1}^{m} \left(2p_i - \delta_{ij} - \delta_{il}\right) \delta_{lk} \left(A_l^{+} A_k^{+}\right)^{\frac{1}{4}} \qquad \text{(S13)} \\
&= \frac{1}{\sqrt{K}\left(\sum\limits_{r=1}^{m} x_r\right)^2} \left(\left(2p_i - \delta_{ij}\right) \left(A_k^{+}\right)^{\frac{1}{2}} - \delta_{ik} \left(A_i^{+} A_k^{+}\right)^{\frac{1}{4}}\right) \qquad \text{(S14)} \\
&= \frac{1}{\sqrt{K}\left(\sum\limits_{r=1}^{m} x_r\right)^2} \left(2p_i - \delta_{ij} - \delta_{ik}\right) \left(A_k^{+}\right)^{\frac{1}{2}} \qquad \text{(S15)}
\end{aligned}$$

and thus:

$$\begin{aligned}
\frac{1}{K} \left(\mathbf{D}^{\mathrm{T}} H_{\mathbf{x}} f_i \mathbf{D}\right)_{lk} &= \frac{1}{K} \sum_{j=1}^{m} \left(\mathbf{D}^{\mathrm{T}}\right)_{lj} \left(H_{\mathbf{x}} f_i \mathbf{D}\right)_{jk} \\
&= \frac{1}{K\left(\sum\limits_{r=1}^{m} x_r\right)^2} \sum_{j=1}^{m} \delta_{lj} \left(A_l^{+} A_j^{+}\right)^{\frac{1}{4}} \left(A_k^{+}\right)^{\frac{1}{2}} \left(2p_i - \delta_{ij} - \delta_{ik}\right) \qquad \text{(S16)} \\
&= \frac{1}{K\left(\sum\limits_{r=1}^{m} x_r\right)^2} \left(A_k^{+}\right)^{\frac{1}{2}} \left(2p_i \left(A_l^{+}\right)^{\frac{1}{2}} - \left(A_i^{+}\right)^{\frac{1}{2}} \delta_{il} - \left(A_l^{+}\right)^{\frac{1}{2}} \delta_{ik}\right) \qquad \text{(S17)}
\end{aligned}$$

Using this, we see that the trace of this matrix is given by:

$$
\begin{aligned}
\frac{1}{K}\mathrm{Tr}[\mathbf{D}^{\mathrm{T}}H_{\mathbf{x}}f_i\mathbf{D}] &= \frac{1}{K}\sum_{k=1}^{m}\left(\mathbf{D}^{\mathrm{T}}H_{\mathbf{x}}f_i\mathbf{D}\right)_{kk} \\
&= \frac{1}{K\left(\sum\limits_{r=1}^{m}x_r\right)^2}\sum_{k=1}^{m}\left(2p_i(A_k^+A_k^+)^{\frac{1}{2}} - (A_i^+A_k^+)^{\frac{1}{2}}\delta_{ik} - (A_k^+A_k^+)^{\frac{1}{2}}\delta_{ik}\right) && \text{(S18)} \\
&= \frac{1}{K\left(\sum\limits_{r=1}^{m}x_r\right)^2}\left(2p_i\left(\sum_{k=1}^{m}A_k^+\right) - 2A_i^+\right) && \text{(S19)}
\end{aligned}
$$

and thus, the second term multiplying dt in (S8) is given by:

$$
\frac{1}{2K}\mathrm{Tr}[\mathbf{D}^{\mathrm{T}}H_{\mathbf{x}}f_i\mathbf{D}] = \frac{-1}{K\left(\sum\limits_{r=1}^{m}x_r\right)^2}\left(A_i^+ - p_i\left(\sum_{k=1}^{m}A_k^+\right)\right) \qquad \text{(S20)}
$$

Finally, denoting $d\mathbf{W}_t = [dW_t^{(1)}, dW_t^{(2)}, \ldots, dW_t^{(m)}]^{\mathrm{T}}$ where each $W_t^{(j)}$ is an independent one dimensional Wiener process, we have:

$$
\begin{aligned}
\left(\mathbf{D}d\mathbf{W}_t\right)_j &= \sum_{k=1}^{m}\mathbf{D}_{jk}dW_t^{(k)} \\
&= \sum_{k=1}^{m}\delta_{jk}\left(A_j^+A_k^+\right)^{\frac{1}{4}}dW_t^{(k)} && \text{(S21)} \\
&= \left(A_j^+\right)^{1/2}dW_t^{(j)} && \text{(S22)}
\end{aligned}
$$

Thus, using (S10), we see that the last term on the RHS of (S8) is given by:

$$
\begin{aligned}
\frac{1}{\sqrt{K}}\left(\nabla_{\mathbf{x}}f\right)^{\mathrm{T}}\mathbf{D}d\mathbf{W}_t &= \frac{1}{\sqrt{K}}\sum_{j=1}^{m}\left(\nabla_{\mathbf{x}}f_i\right)_j\left(\mathbf{D}d\mathbf{W}_t\right)_j \\
&= \frac{1}{\left(\sum\limits_{r=1}^{m}x_r\right)\sqrt{K}}\sum_{j=1}^{m}\left(\delta_{ij} - p_i\right)\left(A_j^+\right)^{1/2}dW_t^{(j)} && \text{(S23)} \\
&= \frac{1}{\left(\sum\limits_{r=1}^{m}x_r\right)\sqrt{K}}\left(A_i^+\right)^{1/2}dW_t^{(i)} - p_i\sum_{j=1}^{m}\left(A_j^+\right)^{1/2}dW_t^{(j)} && \text{(S24)}
\end{aligned}
$$

Putting equations (S12), (S20) and (S24) into (S8) and letting $N_K(t) = \sum_{r=1}^{m} x_r$ we see that $p_i = f_i(\mathbf{X})_t$, the frequency of the i^{th} type in the population $\mathbf{X}_t$, changes according to the equation:

$$\begin{aligned} dp_i = {} & \underbrace{\frac{1}{N_K(t)}\left(A_i^- - p_i\sum_{j=1}^{m}A_j^-\right)dt}_{K\to\infty\text{ prediction}} - \frac{1}{K}\underbrace{\frac{1}{N_K^2(t)}\left(A_i^+ - p_i\left(\sum_{k=1}^{m}A_k^+\right)\right)dt}_{\substack{\text{Directional finite size effects}\\ \text{due to differential turnover rates}}} \\ & + \frac{1}{\sqrt{K}N_K(t)}\underbrace{\left[\left(A_i^+\right)^{1/2}dW_t^{(i)} - p_i\sum_{j=1}^{m}\left(A_j^+\right)^{1/2}dW_t^{(j)}\right]}_{\substack{\text{Non-directional finite size effects}\\ \text{due to stochastic fluctuations}}} \end{aligned} \tag{S25}$$

Plugging the functional forms of (2) and the definitions of w_i and τ_i into the definitions of A_i^- and A_i^+, we obtain the relations

$$\begin{aligned} A_i^- &= x_i w_i(\mathbf{x}) + \lambda Q_i(\mathbf{x}) \\ A_i^+ &= x_i \tau_i(\mathbf{x}) + \lambda Q_i(\mathbf{x}) \end{aligned} \tag{S26}$$

Thus, for the first term of (S25), we have

$$\begin{aligned} \frac{1}{N_K(t)}\left(A_i^- - p_i\sum_{j=1}^{m}A_j^-\right) &= \frac{1}{N_K(t)}\left[w_i(\mathbf{x})x_i + \lambda Q_i(\mathbf{x})\right] - \frac{p_i}{N_K(t)}\sum_{j=1}^{m}\left[w_j(\mathbf{x})x_j + \lambda Q_j(\mathbf{x})\right] \\ &= w_i(\mathbf{x})p_i + \frac{\lambda}{N_K(t)}Q_i(\mathbf{x}) - p_i\sum_{j=1}^{m}\left[w_j(\mathbf{x})p_j + \frac{\lambda}{N_K(t)}Q_j(\mathbf{x})\right] \end{aligned}$$

where we have used the definition of p_i from (5). Now using the definition of mean fitness from (6) and rearranging terms gives us

$$\frac{1}{N_K(t)}\left(A_i^- - p_i\sum_{j=1}^{m}A_j^-\right) = \left(w_i(\mathbf{x}) - \overline{w}\right)p_i + \lambda\left[Q_i(\mathbf{p}) - p_i\left(\sum_{j=1}^{m}Q_j(\mathbf{p})\right)\right] \tag{S27}$$

where we have defined $Q_j(\mathbf{p}) = Q_j(\mathbf{x})/N_K(t)$. Repeating the exact same calculations for the A_i^+ terms in the second term of (S25) now yields equation

$$
\begin{aligned}
dp_i(t) = & \underbrace{\left[(w_i(\mathbf{x}) - \overline{w})p_i + \lambda \left\{ Q_i(\mathbf{p}) - p_i \left(\sum_{j=1}^{m} Q_j(\mathbf{p}) \right) \right\} \right] dt}_{\substack{\text{Infinite population predictions: selection-mutation balance} \\ \text{for higher fitness}}} \\
& - \frac{1}{K} \underbrace{\frac{1}{N_K(t)} \left[(\tau_i(\mathbf{x}) - \overline{\tau})p_i + \lambda \left\{ Q_i(\mathbf{p}) - p_i \left(\sum_{j=1}^{m} Q_j(\mathbf{p}) \right) \right\} \right] dt}_{\substack{\text{Directional noise-induced effects: selection-mutation balance} \\ \text{for lower turnover rates}}} \\
& + \frac{1}{\sqrt{K} N_K(t)} \underbrace{\left[\sqrt{A_i^{+}(\mathbf{x})} dW_t^{(i)} - p_i \sum_{j=1}^{m} \sqrt{A_j^{+}(\mathbf{x})} dW_t^{(j)} \right]}_{\substack{\text{Non-directional noise-induced effects} \\ \text{due to stochastic fluctuations}}}
\end{aligned}
\tag{S28}
$$

which is the first key result (10) presented in the main text (with $\lambda = 0$).

S3 A stochastic analog of the Price equation for finite, fluctuating populations

In this section, we will derive an SDE for the rate of change of the population mean value of any type-level quantity in finite, fluctuating populations. Let f be any type-level quantity, with value $f_i(t)$ for the i^{th} type. Using the product rule of calculus on the definition (6) of the statistical mean tells us that we have the relation

$$
\frac{d\overline{f}}{dt} = \frac{d}{dt}\left(\sum_{i=1}^{m} f_i p_i \right) = \sum_{i=1}^{m} \left(f_i \frac{\partial p_i}{\partial t} + p_i \frac{\partial f_i}{\partial t} \right) = \sum_{i=1}^{m} f_i \frac{\partial p_i}{\partial t} + \overline{\left(\frac{\partial f}{\partial t} \right)} \tag{S29}
$$

i.e.

$$
d\overline{f} = \sum_{i=1}^{m} f_i dp_i + \overline{\left(\frac{\partial f}{\partial t} \right)} dt \tag{S30}
$$

We will further simplify the first term on the RHS of (S30). We do this by using (10), which gives us a representation of dp_i. Using the RHS of (10), we can conclude that we must have

$$\begin{aligned}\sum_{i=1}^{m} f_i dp_i = &\left(\sum_{i=1}^{m} f_i w_i(\mathbf{x}) p_i - \overline{w}\sum_{i=1}^{m} f_i p_i + \lambda\left[\sum_{i=1}^{m} Q_i(\mathbf{p}) f_i - \left(\sum_{j=1}^{m} Q_j(\mathbf{p})\sum_{i=1}^{m} p_i f_i\right)\right]\right) dt \\ &- \frac{1}{KN_K}\left(\sum_{i=1}^{m} f_i \tau_i(\mathbf{x}) p_i - \overline{\tau}\sum_{i=1}^{m} f_i p_i + \lambda\left[\sum_{i=1}^{m} Q_i(\mathbf{p}) f_i - \left(\sum_{j=1}^{m} Q_j(\mathbf{p})\sum_{i=1}^{m} p_i f_i\right)\right]\right) dt \\ &\frac{1}{\sqrt{K}N_K}\left(\left[\sum_{i=1}^{m} f_i\sqrt{A_i^{+}(\mathbf{x})}dW_t^{(i)} - \sum_{i=1}^{m} f_i p_i\sum_{j=1}^{m}\sqrt{A_j^{+}(\mathbf{x})}dW_t^{(j)}\right]\right)\end{aligned} \tag{S31}$$

now using the definition of the statistical mean from (6) in equation (S31), we obtain

$$\begin{aligned}\sum_{i=1}^{m} f_i dp_i = &\left(\overline{fw} - \overline{f}\,\overline{w} + \lambda\left[\sum_{i=1}^{m} Q_i(\mathbf{p}) f_i - \left(\sum_{j=1}^{m} Q_j(\mathbf{p})\overline{f}\right)\right]\right) dt \\ &- \frac{1}{KN_K}\left(\overline{f\tau} - \overline{f}\,\overline{\tau} + \lambda\left[\sum_{i=1}^{m} Q_i(\mathbf{p}) f_i - \left(\sum_{j=1}^{m} Q_j(\mathbf{p})\overline{f}\right)\right]\right) dt \\ &\frac{1}{\sqrt{K}N_K}\left(\left[\sum_{i=1}^{m} f_i\sqrt{A_i^{+}(\mathbf{x})}dW_t^{(i)} - \sum_{j=1}^{m}\overline{f}\sqrt{A_j^{+}(\mathbf{x})}dW_t^{(j)}\right]\right)\end{aligned} \tag{S32}$$

By the definition of the statistical covariance (7), we now obtain

$$\begin{aligned}\sum_{i=1}^{m} f_i dp_i = &\operatorname{Cov}(w,f)dt + \lambda\left[\sum_{i=1}^{m} Q_i(\mathbf{p}) f_i - \overline{f}\left(\sum_{j=1}^{m} Q_j(\mathbf{p})\right)\right] dt \\ &- \frac{1}{KN_K}\left(\operatorname{Cov}(w,f)dt + \lambda\left[\sum_{i=1}^{m} Q_i(\mathbf{p}) f_i - \overline{f}\left(\sum_{j=1}^{m} Q_j(\mathbf{p})\right)\right] dt\right) \\ &\frac{1}{\sqrt{K}N_K}\left(\sum_{i=1}^{m}\left(f_i - \overline{f}\right)\sqrt{A_i^{+}(\mathbf{x})}dW_t^{(i)}\right)\end{aligned} \tag{S33}$$

Collecting all terms that capture effects related to mutations/migrations (*i.e.* all terms with a λ factor) via defining the term

$$M_{\overline{f}}(\mathbf{p}, N_K) := \lambda\left(1 - \frac{1}{KN_K(t)}\right)\left(\sum_{i=1}^{m} f_i Q_i(\mathbf{p}) - \overline{f}\sum_{i=1}^{m} Q_i(\mathbf{p})\right) \tag{S34}$$

and collecting all stochastic integral terms via defining the term

$$dW_{\overline{f}} := \sum_{i=1}^{m}\left(f_i - \overline{f}\right)\sqrt{A_i^{+}(\mathbf{x})}dW_t^{(i)} \tag{S35}$$

and substituting into equation (S33) now yields

$$\sum_{i=1}^{m} f_i dp_i = \mathrm{Cov}(w,f)dt - \frac{1}{KN_K(t)}\mathrm{Cov}(\tau,f)dt + M_{\overline{f}}(\mathbf{p},N_K)dt + \frac{1}{\sqrt{K}N_K(t)}dW_{\overline{f}} \tag{S36}$$

This is the simplified version of the first term on the RHS of equation (S30). Upon substitution, (S30) becomes

$$d\overline{f} = \mathrm{Cov}(w,f)dt - \frac{1}{KN_K(t)}\mathrm{Cov}(\tau,f)dt + M_{\overline{f}}(\mathbf{p},N_K)dt + \frac{1}{\sqrt{K}N_K(t)}dW_{\overline{f}} + \overline{\left(\frac{\partial f}{\partial t}\right)}dt \tag{S37}$$

which is precisely equation (26) in the main text once we set $\lambda = 0$ (i.e. $M_{\overline{f}} = 0$).

S4 A Price-like equation for the variance of a type-level quantity

In this section, we will derive an SDE for the rate of change of the variance of any type-level quantity in finite, fluctuating populations. From the definition (7), we see that the variance of any type level quantity f is given by:

$$\sigma_f^2 := \mathrm{Cov}(f,f) = \overline{(f^2)} - (\overline{f})^2 \tag{S38}$$

By the product rule, we have

$$\frac{d\sigma_f^2}{dt} = 2\overline{f\frac{\partial f}{\partial t}} + \sum_{i=1}^{m} f_i^2 \frac{dp_i}{dt} - \frac{d}{dt}(\overline{f}^2) \tag{S39}$$

We will evaluate the RHS term by term. The first term is as simplified as can be without more information about f. For the second term, we can substitute dp_i from (10) and then use the exact same steps we carried out in supplementary section S3 to derive equation (26). Upon doing this, we obtain

$$\begin{aligned}\sum_{i=1}^{m} f_i^2 dp_i = {} & \mathrm{Cov}(w,f^2)dt - \frac{1}{KN_K}\mathrm{Cov}(\tau,f^2)dt \\ & + \lambda\left(1 - \frac{1}{KN_K(t)}\right)\left(\sum_{i=1}^{m} f_i^2 Q_i(\mathbf{p}) - \overline{f^2}\sum_{i=1}^{m} Q_i(\mathbf{p})\right)dt \\ & + \frac{1}{\sqrt{K}N_K(t)}\left(\sum_{i=1}^{m}\left(f_i^2 - \overline{f^2}\right)\sqrt{A_i^{+}}dW_t^{(i)}\right)\end{aligned} \tag{S40}$$

For the third term, we require Itô's formula. Here, the relevant version of Itô's formula is the one-dimensional version of (S8). Given a one-dimensional process $dX_t = S(X_t)dt + \sum_j D_j(X_t)dW_t^{(j)}$ with S, D_j being suitable real functions and each $W_t^{(j)}$ being an independent Wiener process, Itô's formula says that given any $C^2(\mathbb{R})$ function $g(x)$, we have the relation:

$$dg(X_t) = \left(S(X_t)g'(X_t) + \frac{g''(X_t)}{2}\sum_j D_j^2(X_t)\right)dt + \sum_j D_j(X_t)g'(X_t)dW_t^{(j)} \tag{S41}$$

In our case, we have a one-dimensional process for the mean value $d\overline{f}$ of the type level quantity, and the $C^2(\mathbb{R})$ function $g(x) = x^2$. Itô’s formula thus says that the third term of (S39) is given by:

$$d(\overline{f}^2) = \left(2\overline{f}S(X_t) + \sum_j D_j^2(X_t)\right)dt + \sum_j 2\overline{f}D_j(X_t)dW_t^{(j)} \tag{S42}$$

where the relevant functions S and D_j can be read off from (26). Since the $dW_t^{(j)}$ terms are unwieldy, we will denote the contribution of all the $dW_t^{(j)}$ terms collectively by $dW_{\sigma_f^2}$ to reduce notational clutter and only explicitly calculate these terms at the end. We also note that the covariance operator is a bilinear form, *i.e.* given any three quantities X, Y and Z and any constant $a \neq 0$, we have the relations:

$$\text{Cov}(aX, Y) = a\text{Cov}(X, Y) = \text{Cov}(X, aY)$$

$$\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$$

Substituting equations (S40) and (S42) into equation (S39) and using this property of covariances, we obtain:

$$\begin{aligned} d\sigma_f^2 = {} & \text{Cov}(w, f^2 - 2\overline{f}f)dt - \frac{1}{KN_K}\left(\text{Cov}(\tau, f^2 - 2\overline{f}f)\right)dt + 2\left(\overline{f\frac{\partial f}{\partial t}} - \overline{f}\,\overline{\left(\frac{\partial f}{\partial t}\right)}\right)dt \\ & + \lambda\left(1 - \frac{1}{KN_K(t)}\right)\left(\sum_{i=1}^{m}(f_i^2 - 2\overline{f}f_i)Q_i(\mathbf{p}) - (\overline{f^2} - 2\overline{f}^2)\sum_{i=1}^{m}Q_i(\mathbf{p})\right)dt \\ & - \frac{1}{KN_K^2(t)}\left(\sum_{i=1}^{m}(f_i - \overline{f})^2 A_i^+\right)dt \\ & + dW_{\sigma_f^2} \end{aligned} \tag{S43}$$

Now, we note that

$$\frac{1}{N_K}A_i^+ = \frac{1}{N_K}\left(\tau_i x_i + \lambda Q_i(\mathbf{x})\right) \tag{S44}$$

$$= \tau_i p_i + \lambda Q_i(\mathbf{p}) \tag{S45}$$

and thus the third line of (S43) is

$$\frac{1}{KN_K^2(t)}\left(\sum_{i=1}^{m}(f_i-\overline{f})^2 A_i^+\right)dt = \frac{1}{KN_K}\sum_{i=1}^{m}(f_i-\overline{f})^2\left(\tau_i p_i + \lambda Q_i(\mathbf{p})\right) \tag{S46}$$

$$= \frac{1}{KN_K}\sum_{i=1}^{m}\left(f_i-\overline{f}\right)^2\left(\tau_i p_i + \lambda Q_i(\mathbf{p})\right) \tag{S47}$$

$$= \frac{1}{KN_K}\left(\overline{\tau\left(f-\overline{f}\right)^2} + \lambda\sum_{i=1}^{m}\left(f_i-\overline{f}\right)^2 Q_i(\mathbf{p})\right) \tag{S48}$$

$$= \frac{1}{KN_K}\left(\text{Cov}(\tau, \left(f-\overline{f}\right)^2) + \overline{\tau}\,\overline{\left(f-\overline{f}\right)^2} + \lambda\sum_{i=1}^{m}\left(f_i-\overline{f}\right)^2 Q_i(\mathbf{p})\right) \tag{S49}$$

$$= \frac{1}{KN_K}\left(\text{Cov}(\tau, \left(f-\overline{f}\right)^2) + \overline{\tau}\sigma_f^2 + \lambda\sum_{i=1}^{m}\left(f_i-\overline{f}\right)^2 Q_i(\mathbf{p})\right) \tag{S50}$$

where we have used the definition of statistical covariance in the second to last line and used the definition of statistical variance in the last line. Substituting (S50) into (S43) and using $M_{\sigma_f^2}(\mathbf{p}, N_K)$ to denote the contributions of all the mutational terms (*i.e.* all terms with a λ factor) for notational brevity, we obtain

$$\begin{aligned} d\sigma_f^2 = {} & \text{Cov}(w, f^2 - 2\overline{f}f)dt - \frac{1}{KN_K}\left(\text{Cov}(\tau, f^2 - 2\overline{f}f) + \text{Cov}(\tau, \left(f-\overline{f}\right)^2) + \overline{\tau}\sigma_f^2\right)dt \\ & + 2\text{Cov}\left(\frac{\partial f}{\partial t}, f\right)dt + M_{\sigma_f^2}(\mathbf{p}, N_K)dt + dW_{\sigma_f^2} \end{aligned} \tag{S51}$$

We can now complete the square inside the covariance terms of the first line of the RHS by writing $f^2 - 2\overline{f}f = (f-\overline{f})^2 - \overline{f}^2$ to obtain

$$\begin{aligned} d\sigma_f^2 &= \left[\operatorname{Cov}\left(w, (f-\overline{f})^2\right) - \operatorname{Cov}\left(w, \left(\overline{f}\right)^2\right) \right] dt \\ &\quad - \frac{1}{KN_K}\left[\operatorname{Cov}\left(\tau, (f-\overline{f})^2\right) - \operatorname{Cov}\left(\tau, \left(\overline{f}\right)^2\right) + \operatorname{Cov}(\tau, \left(f-\overline{f}\right)^2) + \overline{\tau}\sigma_f^2 \right] dt \\ &\quad + 2\operatorname{Cov}\left(\frac{\partial f}{\partial t}, f\right) dt + M_{\sigma_f^2}(\mathbf{p}, N_K)dt + dW_{\sigma_f^2} \end{aligned} \tag{S52}$$

To simplify the covariance terms of the first line of the RHS, we observe that

$$\begin{aligned} \operatorname{Cov}\left(w, \left(\overline{f}\right)^2\right) &= \overline{\left(w\left(\overline{f}\right)^2\right)} - \overline{w}\,\overline{\left(\left(\overline{f}\right)^2\right)} \\ &= \left(\overline{f}\right)^2 \sum_{i=1}^{m} w_i p_i - \overline{w}\left(\overline{f}\right)^2 \sum_{i=1}^{m} p_i \\ &= \left(\overline{f}\right)^2 \overline{w} - \overline{w}\left(\overline{f}\right)^2 = 0 \end{aligned}$$

and similarly,

$$\operatorname{Cov}\left(\tau, \left(\overline{f}\right)^2\right) = 0$$

and thus, using this in (S52), we see that the rate of change of the variance of any type-level quantity f in the population satisfies:

$$\begin{aligned} d\sigma_f^2 &= \operatorname{Cov}\left(w, (f-\overline{f})^2\right) dt - \frac{1}{KN_K}\left[\overline{\tau}\sigma_f^2 + 2\operatorname{Cov}\left(\tau, (f-\overline{f})^2\right) \right] dt \\ &\quad + 2\operatorname{Cov}\left(\frac{\partial f}{\partial t}, f\right) dt + M_{\sigma_f^2}(\mathbf{p}, N_K)dt + dW_{\sigma_f^2} \end{aligned} \tag{S53}$$

This is precisely equation (31) in the main text. To calculate the mutation term, we substitute (S50) into (S43) to find

$$\begin{aligned} M_{\sigma_f^2}(\mathbf{p}, N_K) &= \lambda\left(\sum_{i=1}^{m} \left(f_i^2 - 2\overline{f} f_i - \overline{f^2} + 2\overline{f}^2 \right) Q_i(\mathbf{p}) \right) \\ &\quad - \frac{\lambda}{KN_K} \sum_{i=1}^{m} \left(f_i^2 - 2\overline{f} f_i - \overline{f^2} + 2\overline{f}^2 + (f_i - \overline{f})^2 \right) Q_i(\mathbf{p}) \end{aligned} \tag{S54}$$

We can further simplify the first term of the RHS as

$$\begin{aligned} f_i^2 - 2\overline{f} f_i - \overline{f^2} + 2\overline{f}^2 &= (f_i^2 + \overline{f}^2 - 2\overline{f} f_i) - (\overline{f^2} - \overline{f}^2) \\ &= (f_i - \overline{f})^2 - \sigma_f^2 \end{aligned}$$

and similarly, the second term as

$$f_i^2 - 2\overline{f} f_i - \overline{f^2} + 2\overline{f}^2 + (f_i - \overline{f})^2 = 2(f_i - \overline{f})^2 - \sigma_f^2$$

thus, the contributions of influx terms to the change in the variance of f are given by

$$\begin{aligned} M_{\sigma_f^2}(\mathbf{p}, N_K) = \lambda \left(\sum_{i=1}^{m} \left((f_i - \overline{f})^2 - \sigma_f^2 \right) Q_i(\mathbf{p}) \right) \\ - \frac{\lambda}{K N_K} \sum_{i=1}^{m} \left(2(f_i - \overline{f})^2 - \sigma_f^2 \right) Q_i(\mathbf{p}) \end{aligned} \tag{S55}$$

which after slight rearrangement becomes

$$M_{\sigma_f^2}(\mathbf{p}, N_K) = \lambda \left(\sum_{i=1}^{m} \left[\left(1 - \frac{2}{K N_K} \right) (f_i - \overline{f})^2 Q_i(\mathbf{p}) \right] - \sigma_f^2 \left(1 - \frac{1}{K N_K} \right) \sum_{i=1}^{m} Q_i(\mathbf{p}) \right) \tag{S56}$$

Finally, for the stochastic integral term, we can use equations (S40) and (S42) to calculate:

$$dW_{\sigma_f^2} = \frac{1}{\sqrt{K} N_K(t)} \left(\sum_{i=1}^{m} \left(f_i^2 - \overline{f^2} - 2\overline{f}(f_i - \overline{f}) \right) \sqrt{A_i^+} dW_t^{(i)} \right) \tag{S57}$$

$$= \frac{1}{\sqrt{K} N_K(t)} \left(\sum_{i=1}^{m} \left(f_i^2 - \overline{f^2} - 2\overline{f} f_i - 2\overline{f}^2 \right) \sqrt{A_i^+} dW_t^{(i)} \right) \tag{S58}$$

$$= \frac{1}{\sqrt{K} N_K(t)} \left(\sum_{i=1}^{m} \left(f_i - \overline{f} \right)^2 \sqrt{A_i^+} dW_t^{(i)} \right) \tag{S59}$$

which is equation (32) in the main text upon setting $\lambda = 0$ (*i.e.* $M_{\sigma^2} = 0$).

S5 A more elegant representation of sums of stochastic integrals against independent Wiener processes

In the main text, we arrived at three stochastic differential equations (equations (10), (26), and (31)) that describe the change in the frequency of a type, the population mean value of a type-level quantity, and the population variance of a type-level quantity over time. All three of these equations contained stochastic fluctuation terms which were of the form of a sum of stochastic integrals of several independent functions against independent Wiener processes. In this section, we will present a more elegant representation of these terms.

Let us first recall that given m independent one-dimensional Wiener processes $W_t^{(1)}, W_t^{(2)}, \dots, W_t^{(m)}$, m ‘nice’ real functions $g_1(x), g_2(x), \dots, g_m(x)$, and the stochastic process

$$dX_t = \sum_{i=1}^{m} g_i(X_t)\,dW_t^{(i)}$$

We can always find a *single* one-dimensional Wiener process W_t such that

$$dX_t = \left(\sum_{i=1}^{m} g_i^2(X_t)\right)^{1/2} dW_t$$

This result is well-known but we were unable to find a reference that explicitly proved it, and so we prove it as a lemma at the end of this supplementary section.

Using this result, we can now calculate the stochastic integral terms of our equations. For equation (12), we can calculate

$$\sum_{i=1}^{m} \left(x_i\tau_i + \lambda Q_i(\mathbf{x})\right) dW_t^{(i)} = \left[\sum_{i=1}^{m} x_i\tau_i + \lambda \sum_{i=1}^{m} Q_i(\mathbf{x})\right]^{1/2} dW_t^{N_K} \tag{S60}$$

$$= \left[\overline{\tau} N_K(t) + \lambda \sum_{i=1}^{m} Q_i(\mathbf{x})\right]^{1/2} dW_t^{N_K} \tag{S61}$$

where $W_t^{N_K}$ is a one-dimensional Wiener process. For equation (S28), the stochastic analog of the replicator-mutator equation, we find that the noise term can be written as a stochastic integral against a single Wiener process W_t as

$$\frac{1}{\sqrt{KN_K(t)}} \left[p_i(1-p_i)^2\tau_i + p_i^2\left(\sum_{j\neq i}\tau_j p_j\right) + \lambda\left\{(1-p_i)^2 Q_i(\mathbf{p}) + p_i^2 \sum_{j\neq i} Q_j(\mathbf{p})\right\}\right]^{1/2} dW_t \tag{S62}$$

For equation (S37), the stochastic analog of the Price equation, we have:

$$dW_{\overline{f}} = \sum_{i=1}^{m}\left(f_i - \overline{f}\right)\sqrt{A_i^+(\mathbf{x})}\,dW_t^{(i)} = \left(\sum_{i=1}^{m}\left(f_i - \overline{f}\right)^2 A_i^+(\mathbf{x})\right)^{1/2} dW_t \tag{S63}$$

where W_t is now a single one-dimensional Wiener process. This is precisely the term calculated in equation (S50) (barring the $1/KN_K^2$ pre-factor), and thus the stochastic term for the mean value is given by:

$$dW_{\overline{f}} = \sqrt{N_K(t)\left(\mathrm{Cov}(\tau, \left(f-\overline{f}\right)^2) + \overline{\tau}\sigma_f^2 + \lambda \sum_{i=1}^{m} \left(f_i - \overline{f}\right)^2 Q_i(\mathbf{p})\right)}\, dW_t \tag{S64}$$

Similarly, for the variance equation (S57), we can write

$$dW_{\sigma_f^2} = \sum_{i=1}^{m} \left(f_i - \overline{f}\right)^2 \sqrt{A_i^+(\mathbf{x})} dW_t^{(i)} = \left(\sum_{i=1}^{m} \left(f_i - \overline{f}\right)^4 A_i^+(\mathbf{x})\right)^{1/2} dW_t \tag{S65}$$

where W_t is now a single one-dimensional Wiener process. A calculation exactly analogous to that done in obtaining (S50) reveals that this term can be written as

$$dW_{\sigma_f^2} = \sqrt{N_K(t)\left(\mathrm{Cov}(\tau, \left(f-\overline{f}\right)^4) + \overline{\tau}(\sigma_f^2)^2 + \lambda \sum_{i=1}^{m} \left(f_i - \overline{f}\right)^4 Q_i(\mathbf{p})\right)}\, dW_t \tag{S66}$$

Proof of the representation of sums of stochastic integrals with respect to independent Wiener processes

Here, we prove the mathematical result we used above. We stress once again that this is not a new result — we provide the proof here because, while the proof is mathematically easy, we were unable to find a suitable citation that explicitly writes down the proof.

Lemma. Let $m \in \mathbb{N}$. Let $W_t^{(1)}, W_t^{(2)}, \ldots, W_t^{(m)}$ be m independent one-dimensional Wiener processes. Let $g_1(x), g_2(x), \ldots, g_m(x)$ be m 'nice' ($L^2(\mathbb{R})$, Lipschitz, *etc.*) real functions. Let

$$dX_t = \sum_{i=1}^{m} g_i(X_t) dW_t^{(i)}$$

Then, we can always find a *single* one-dimensional Wiener process W_t (on the same probability space) such that

$$dX_t = \left(\sum_{i=1}^{m} g_i^2(X_t)\right)^{1/2} dW_t$$

Proof. It suffices to prove the $m = 2$ case.
Let $dX_t = g_1(X_t)dW_t^{(1)} + g_2(X_t)dW_t^{(2)}$. Let us consider the *two*-dimensional process $\mathbf{W}_t = [W_t^{(1)}, W_t^{(2)}]^{\mathrm{T}}$ on $\mathbb{R}^2$. Define a new function $G : \mathbb{R} \to \mathbb{R}^2$ given by

$$G(x) = \frac{1}{\sqrt{g_1^2(x) + g_2^2(x)}} \begin{bmatrix} g_1(x) \\ g_2(x) \end{bmatrix} \tag{S67}$$

Now, by definition, we have

$$\int_0^t G(X_s)\cdot d\mathbf{W}_s = \int_0^t \frac{g_1(X_s)}{\sqrt{g_1^2(X_s)+g_2^2(X_s)}}dW_t^{(1)} + \int_0^t \frac{g_2(X_s)}{\sqrt{g_1^2(X_s)+g_2^2(X_s)}}dW_s^{(2)} \tag{S68}$$

Using the Itô isometry (Karatzas and Shreve, 1998, Chapter 2, Proposition 2.10), we can calculate the quadratic variation of $\int G\cdot d\mathbf{W}$ as

$$\begin{aligned}\left\langle \int G(X_s)\cdot d\mathbf{W}_s \right\rangle_t &= \int_0^t \left\|G(X_s)\right\|^2 d\langle \mathbf{W}\rangle_s \\ &= \int_0^t \frac{1}{g_1^2+g_2^2}\cdot (g_1^2+g_2^2)ds \\ &= \int_0^t ds = t \end{aligned}\tag{S69}$$

Since $\int G\cdot d\mathbf{W}$ is a stochastic integral, the process $(t,\omega)\to\int_0^t G(X_s(\omega))\cdot d\mathbf{W}_s(\omega)\in\mathscr{M}_2^c$ and is thus a continuous martingale. But, by Lévy's characterization of Brownian motion (Karatzas and Shreve, 1998, Chapter 3, Theorem 3.16), the only continuous martingale M_t that satisfies $\langle M\rangle_t = t$ is the standard Brownian motion. Thus, from equation (S69), we are led to conclude that there is a one-dimensional Wiener process W_t on the same probability space such that we can write

$$G(X_t)\cdot d\mathbf{W}_t = dW_t \tag{S70}$$

We can now use equation (S68) on the LHS of equation (S70) to write

$$\frac{g_1(X_t)}{\sqrt{g_1^2(X_t)+g_2^2(X_t)}}dW_t^{(1)} + \frac{g_2(X_t)}{\sqrt{g_1^2(X_t)+g_2^2(X_t)}}dW_t^{(2)} = dW_t \tag{S71}$$

$$\Rightarrow g_1 dW_t^{(1)} + g_2 dW_t^{(2)} = \sqrt{g_1^2(X_t)+g_2^2(X_t)}dW_t \tag{S72}$$

By definition of our original process X_t, we can now conclude that

$$dX_t = \sqrt{g_1^2(X_t)+g_2^2(X_t)}dW_t \tag{S73}$$

thus completing the proof. □

S6 Long-time behavior of the stochastic replicator equation for two species

To study the effects of demographic stochasticity on evolutionary dynamics more thoroughly, we use this section to examine the long-term behavior of the system defined by equation (13). Following McLeod and Day, 2019, we will do this using the speed density. Given any one-dimensional diffusion process $dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$ defined over an interval $[a, b] \subseteq \mathbb{R}$, the *speed density* $m(x)$ of the process (Karlin and Taylor, 1981; Etheridge, 2011) is defined as the function

$$m(x) = \frac{1}{\sigma^2(x)} \exp\left(2 \int^{x} \frac{\mu(y)}{\sigma^2(y)} dy\right) \tag{S74}$$

where the lower limit of the integral being left unspecified is meant to denote an indefinite integral evaluated at the point x since the choice of the lower limit is arbitrary (Karlin and Taylor, 1981, Chapter 15, Equation 3.10). The speed density is important because it provides information about the long-term behavior of the stochastic process X_t (Karlin and Taylor, 1981, Chapter 15, Remark 3.2). In particular, if there exists a constant $0 < \mathcal{N} < \infty$ such that $\int_a^b m(x)dx = 1/\mathcal{N}$, then the stochastic process obtained as the solution to $dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$ attains a unique stationary state X_∞ as $t \to \infty$, and this stationary state has a probability distribution given by(Karlin and Taylor, 1981, Chapter 15, Equation 5.34 along with Chapter 15, Equation 3.10; Czuppon and Traulsen, 2021)

$$\mathbb{P}\left(\{x_1 \leq X_\infty \leq x_2\}\right) = \mathcal{N} \int_{x_1}^{x_2} m(x)dx + C \tag{S75}$$

That is to say, the probability density of the stationary state will be given by $\mathcal{N}m(x)$. Regardless of whether such an $\mathcal{N}$ can be found, the speed density $m(x)$ always tells us about the time the system spends in the vicinity of the point x. More precisely, if we provide an initial condition $x_0 \in [a, b]$ for the stochastic process obtained as the solution to $dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$, the expected time taken by this process to exit the interval $(x_0 - \epsilon, x_0 + \epsilon)$ is proportional to $m(x_0)$ as $\epsilon \to 0$ (Karlin and Taylor, 1981, Chapter 15, Remark 3.2). In our case, we have a stochastic process for the change of type frequencies over time that takes values in $[0, 1]$ and is given by the solution to equation (14). In the rest of this section, we work with equation (13a) and thus do not account for influx terms λQ_i. For convenience, let us define

$$E(p) = s(p, N_K) + \frac{1}{KN_K}\kappa(p, N_K) \tag{S76}$$

$$V(p) = \frac{1}{KN_K}\left(\tau_1(p, N_K) + p\kappa(p, N_K)\right) \tag{S77}$$

where we have suppressed the N_K dependence of E and V to reduce clutter. In this notation, equation (14) becomes

$$dp = p(1-p)E(p)dt + \sqrt{p(1-p)V(p)}dW_t \tag{S78}$$

Comparing terms with (S74), we see that the speed density of our process is given by

$$m(p) = \frac{1}{p(1-p)V(p)} \exp\left(2\int^{p} \frac{E(q)}{V(q)}dq\right) \tag{S79}$$

For general functions $E(p)$ and $V(p)$, it is very often impossible to analytically calculate or predict the behavior of the complete function defined by (S79). However, since we are primarily interested in which trait frequencies p are likely, we can still make analytical progress by examining the derivative dm/dp. If dm/dp is a strictly increasing function of p, then higher values of frequency p are always favored, and species 2 is expected to go extinct more often than species 1. Likewise, if dm/dp is a strictly decreasing function of p, lower frequencies of p are favored, and species 1 is expected to go extinct. Lastly, points at which $dm/dp = 0$ correspond to extrema of the speed density and can thus be used to find the most likely and least likely values of trait frequency in the system.

We would therefore like to examine the behavior of dm/dp as a function of p. Differentiating both sides of equation (S79) with respect to p, we find

$$\frac{dm}{dp} = m(p)\left[\frac{2p-1}{p(1-p)} + 2\frac{E(p)}{V(p)} - \frac{1}{V(p)}\frac{dV}{dp}\right] \tag{S80}$$

which is Eq. 15 in the main text.

After substituting the functional form of $V(p)$ from equation (S77), this yields (after some lines of algebra):

$$\frac{dm}{dp} = m(p)\left[\frac{2p-1}{p(1-p)} + 2\frac{E(p)}{V(p)} + \frac{1}{N_K}\frac{dN_K}{dp} - \frac{1}{V(p)}\left(\kappa(p, N_K) + (1-p)\frac{d\tau_1}{dp} + p\frac{d\tau_2}{dp}\right)\right] \tag{S81}$$

Let us examine each term on the RHS of equation (S81). The first term on the RHS is $(2p-1)/p(1-p)$. This expression is (anti)-symmetric about $p = 0.5$ and always drives the system towards the boundaries of $[0,1]$. It is thus uninteresting for calculating the sign of dm/dp.

Since $V(p)$ must clearly be non-negative in order for equation (S78) to be well-defined, the second term, $E(p)/V(p)$, always has the same sign as $E(p)$. Equation (S81) tells us that the speed density (and thus the stationary distribution, when it exists) also depends on contributions from the dW_t term of equation (S78). We have split this contribution into two separate terms, the third and fourth terms on the RHS of equation (S81), each of which we will examine individually.

The third term on the RHS of (S81) captures the effect of the frequency of species 1 on the per-capita growth rate of the population as a whole. Thus, if species 1 is altruistic, mutualistic, or commensal, then dN_K/dp will be positive, whereas if the species is spiteful, competitive, or amensal, dN_K/dp will be negative. The sign of the third term on the RHS of (S81) thus depends on the nature of the ecological interactions that species 1 is involved in — species that increase the per-capita growth rate of the total population are favored, and those that decrease the per-capita growth rate of the total population are disfavored.

The fourth term on the RHS of equation (S81) captures the effects of noise-induced selection acting on differential turnover rates. Since $E(p)$ also has both a $1/V(p)$ factor and a noise-induced selection term, we are better off substituting the functional form of $E(p)$ from (S76) into equation (S81) and collecting all terms with a $1/V(p)$ factor so as to collect all terms corresponding to selection (both classical and noise-induced). Upon doing this, we obtain

$$\frac{dm}{dp} = m(p)\left[\frac{2p-1}{p(1-p)} + \frac{1}{N_K}\frac{dN_K}{dp} + \frac{1}{V(p)}\left(2KN_K s(p,N_K) + \kappa(p,N_K) - (1-p)\frac{d\tau_1}{dp} - p\frac{d\tau_2}{dp}\right)\right] \tag{S82}$$

The interpretations of the first two terms on the RHS of (S82) have already been explained above. Since $V(p)$ is always non-negative, we only need to look at the sign of the expression

$$2KN_K s(p,N_K) + \kappa(p,N_K) - (1-p)\frac{d\tau_1}{dp} - p\frac{d\tau_2}{dp} \tag{S83}$$

The first term of (S83) is the effect of classical selection and has the same sign as the selection coefficient $s(p,N_K)$. Notice that since this term is $\mathcal{O}(K)$ whereas all other terms in equation (S82) are $\mathcal{O}(1)$, this term dominates the dynamics when K is large, again indicating that the effects of natural selection dominate in large populations with non-zero selection coefficient. If instead $Ks(p,N_K)$ is small, either through a small population size, weak selection (or no selection), or both, the other terms of (S83) play a stronger role. The second term of (S83) is simply the noise-induced selection coefficient $\kappa(p)$, and is thus positive whenever $\tau_1 < \tau_2$. This term thus causes the probability density function to be biased towards the species with lower per-capita turnover rates. The last two terms of (S83) capture the frequency-dependence of noise-induced selection.

S7 The infinite population limit recovers standard equations of population biology

In this section, we show how our SDEs recover several classic equations of population biology in the infinite population size limit.

Replicator-mutator equation

If w take $K \to \infty$ in (10), we obtain an ODE that reads:

$$\frac{dp_i}{dt} = (w_i(\mathbf{x}) - \overline{w})p_i + \lambda \left[Q_i(\mathbf{p}) - p_i \left(\sum_{j=1}^{m} Q_j(\mathbf{p}) \right) \right] \tag{S84}$$

The first term of (S84) describes changes due to faithful (non-mutational) replication, and the second describes changes due to mutation. For this reason, equation (S84) is called the *replicator-mutator equation* in the evolutionary game theory literature, where the individual 'types' are interpreted to be pure strategies and the influx rate λ is a mutation rate, denoted by μ. If in addition, each $w_i(\mathbf{x})$ is linear in $\mathbf{x}$, meaning we can write $w_i(\mathbf{x}) = \sum_j a_{ij} x_j$ for some set of constants a_{ij}, then we get the replicator-mutator equation for matrix games, and the constants a_{ij} form the 'payoff matrix'. As is well-known, the replicator equation (without mutation) for matrix games with m pure strategies is equivalent to the generalized Lotka-Volterra equations for a community with $m-1$ species (Hofbauer and Sigmund, 1998), providing the connection to community ecology. Equation (S84) is also equivalent to Eigen's *quasispecies equation* from molecular evolution if each 'type' is interpreted as a genetic sequence and each $w_i(\mathbf{x})$ is a constant function (Page and Nowak, 2002).

(Dynamical) Price equation

Taking $K \to \infty$ in equation (26) recovers the Price equation as the infinite population limit. Here, we mean the Price equation as formulated in continuous time with time-varying characters (Lion, 2018; Day et al., 2020).

$$\frac{d\overline{f}}{dt} = \mathrm{Cov}(w, f) + \overline{\left(\frac{\partial f}{\partial t}\right)} + \lambda \left(\sum_{i=1}^{m} f_i Q_i(\mathbf{p}) - \overline{f} \sum_{i=1}^{m} Q_i(\mathbf{p}) \right) \tag{S85}$$

Many authors additionally assume that the quantity f does not itself change over time at the type level, meaning that $\partial f_i / \partial t \equiv 0 \ \forall \ i$ and the feedback term thus disappears. This yields a somewhat more familiar equation in continuous time (Lion, 2018). Standard texts also usually use a version formulated in discrete time that is more general for single-step changes, but is dynamically insufficient (Frank, 2012; Queller, 2017).

Fisher's fundamental theorem of natural selection

Taking $K \to \infty$ in (28) and noting that the process tends to a deterministic process as $K \to \infty$, as noted in section S7 (and thus the expectation value in the infinite population case is superfluous), we obtain an ODE:

$$\frac{d\overline{w}}{dt} = \sigma_w^2(t) + \overline{\left(\frac{\partial w}{\partial t}\right)} \tag{S86}$$

This is Fisher's fundamental theorem in the presence of ecological feedbacks to fitness (Frank and Slatkin, 1992; Kokko, 2021).

Lion 2018's variance equation

Taking $K \to \infty$ in equation (31) yields

$$\frac{d\sigma_f^2}{dt} = \text{Cov}\left(w, (f - \overline{f})^2\right) + 2\text{Cov}\left(\frac{\partial f}{\partial t}, f\right) + \lambda\left[\left(\sum_{i=1}^{m}(f_i - \overline{f})^2 Q_i(\mathbf{p})\right) + \sigma_f^2 \sum_{i=1}^{m} Q_i(\mathbf{p})\right] \tag{S87}$$

This is precisely equation (14) in Lion, 2018 with influx terms λQ_i.

S8 Recovering some previous studies as special cases

In many social evolution models, cooperators are predicted to go extinct in infinite populations but are actually favored by evolution in finite, fluctuating populations, causing a 'reversal' in the direction of evolution predicted by natural selection (Houchmandzadeh and Vallade, 2012; Chotibut and Nelson, 2015; Constable et al., 2016; McLeod and Day, 2019). McLeod and Day, 2019 have recently shown that such reversals can occur in a wide array of social evolution models due to the same effect that we recognize here as noise-induced selection. Formally, all the models presented in McLeod and Day, 2019 can be recovered in our framework by setting $m = 2$ and $s(\mathbf{x}) = -\epsilon c(\mathbf{x})$ for a constant $\epsilon \in \mathbb{R}$ and a non-negative function $c(\mathbf{x})$ in our stochastic replicator-mutator equation (Eq. 10). The function $T(p)$ in McLeod and Day, 2019 — a quantity that varies in the various models they study — is precisely the mean turnover $\overline{\tau}$ in our framework.

In evolutionary epidemiology, models have shown that reduced virulence is more important than increased transmission rate for pathogen spread in finite, fluctuating populations, especially when the population size is small (Humplik et al., 2014; Parsons et al., 2018; Day et al., 2020). Indeed, if the population is small or selection is weak, slower strains can have higher fixation probabilities than faster strains even if the slower strain has a lower basic reproduction ratio (R_0) than its competitor, causing a complete reversal in the direction of evolution predicted in infinite populations (Parsons et al., 2018). These results have recently been explained in a generic manner using both a replicator-mutator/'stochastic adaptive dynamics' approach (Parsons et al., 2018) and a two-species Price equation formalism (Day et al., 2020), though both these papers use assumptions and language particular to evolutionary epidemiology. We note that equation (2.5) in Parsons et al., 2018 is exactly equivalent to our stochastic replicator-mutator equation with no mutation (equation (10) with $\lambda = 0$) upto a change in notation upon substituting the specific birth and death rate functions chosen in their paper into our equation (10). Similarly, equation (5.1) in Day et al., 2020 is exactly equivalent to our stochastic Price equation for 2 species (equation (26) with $m = 2$) if we write out w and τ in terms of per-capita birth and death rates. Our work can therefore be used to recapitulate these results and show that the effects they illustrate are not particular to epidemiological models.

S9 An exact solution for the example in the main text with equal growth rate when turnover rates vary linearly with frequency

In this section, we provide an exact quasi-stationary distribution for the rate modulation example in which birth and death rates are increased by the same amount (case 1 with $\epsilon_b = \epsilon_d$). In this case, we have

$$E(p) = -\frac{2\epsilon_d}{KN_K} \tag{S88}$$

$$V(p) = \frac{1}{KN_K}(\tau_1(p, N_K) - 2\epsilon_d p) \tag{S89}$$

For notational convenience, let $a = 2\epsilon_d = -\kappa(p)$. To study the system, we need a functional form for τ_1. We assume here (for simplicity) that $\tau_1(p, N_K) = bp + c$, where b and c are constants. c can be viewed as an 'intrinsic' turnover rate, and b as a frequency-dependent component that may be either positive or negative. We are therefore restricting ourselves to linear frequency dependence of τ_1, but allowing both positive and negative frequency-dependence, with the strength of frequency-dependence controlled by $|b|$. Note that since τ_1 is the sum of two rates and $p(1-p)V(p)$ is the infinitesimal variance of the trait frequency SDE, the parameters a, b, and c must be chosen such that $\tau_1(p) = bp + c > 0, V(p) = (b-a)p + c > 0 \ \forall \ p \in [0,1]$ for the system to be biologically meaningful. In particular, $\tau_1(0, N_K)$ and $\tau_1(1, N_K)$ must be non-negative, and we must thus have $\tau_1(0, N_K) = c \geq 0$ and $\tau_1(1, N_K) = b + c > 0$. We must also have $V(1) > 0$, and thus $b + c - a \geq 0$. In our new notation, Eq. S88 and S89 become

$$E(p) = -\frac{a}{KN_K} \tag{S90}$$

$$V(p) = \frac{1}{KN_K}((b-a)p + c) \tag{S91}$$

The speed density of the system can be written (from Eq. S79) as

$$m(p) = \frac{1}{p(1-p)V(p)} \exp\left(2\int^{p} \frac{E(q)}{V(q)}dq\right)$$

$$\Rightarrow m(p) = \frac{KN_K}{p(1-p)((b-a)p+c))} \exp\left(2a\int^{p} \frac{1}{(b-a)q+c}dq\right) \tag{S92}$$

Case 1: No frequency dependence in $V(p)$

If $a = b$, *i.e.* the frequency dependence of τ_1 is positive with strength exactly equal to $2\epsilon_d$, Eq. S92 becomes

$$m(p) = \frac{1}{p(1-p)c} \exp\left(2\int^{p} \frac{-a}{c}dq\right) = C\frac{1}{p(1-p)}e^{-\alpha p} \tag{S93}$$

where $\alpha = 2a/c > 0$ is a positive constant and we use C to denote a constant whose precise value is irrelevant (and thus may change from line to line below — the important thing is that C does not depend on p and thus plays the role of a normalization constant).

The shape of the distribution given by Eq. S93 can be thought of as the combination of two components: The term $p(1-p)$ is symmetric with respect to the transformation $p \to 1-p$ (and thus symmetric about $p = 0.5$) and thus does not favor either type of individual, whereas $e^{-\alpha p}$ is a strictly decreasing function of p and thus always favors lower frequencies of type 1 individuals. If α is very small, the effect of $e^{-\alpha p}$ is negligible and the distribution of types is approximately a symmetric 'U-shaped' parabola centered at 0.5, with $p = 0.5$ being the least likely frequency. This is the expectation we would have if neutral genetic drift was the only force at play: The distribution is (approximately) symmetric with respect to the transformation $p \to 1-p$, with $p = 0$ and $p = 1$ being the most likely states and $p = 0.5$ being the least likely state.

If instead, α is not small, the function $e^{-\alpha p}$ decays quickly and biases the distribution towards lower values of p. In this case, the function is a distorted U-shape, with the minimum point being somewhere in $(1/2, 1)$. The extent of bias towards lower values of p increases as α increases.

Thus, in the case where $2\epsilon_d = d\tau_1/dp$, we can conclude that lower frequencies of type modulators are always more likely in the stationary distribution, and the biasing is stronger as the ratio of the rate modulation (ϵ_d) to the intrinsic frequency-independent turnover rate (c) increases. Note that the shape of the quasi-stationary density (and thus the extent of deviation from neutrality) does *not* depend on the total population size KN_K.

Case 2: Frequency dependence in $V(p)$

Assuming $a \neq b$, we can calculate the exponential term in Eq. S92 as

$$\exp\left(2\int^{p}\frac{E(q)}{V(q)}dq\right) = \exp\left(-2a\int^{p}\frac{1}{(b-a)q+c}dq\right) \tag{S94}$$

$$= C\exp\left(-\frac{2a}{b-a}\left[\log((b-a)p+c)\right]\right) = C[(b-a)p+c]^{-\frac{2a}{b-a}} \tag{S95}$$

where we once again use C to denote a multiplicative constant whose precise value is irrelevant. Thus, the speed density S92 is given by

$$m(p) = \frac{C}{p(1-p)}[(b-a)p+c]^{-(\gamma+1)} \tag{S96}$$

where we have defined $\gamma = 2a/(b-a)$.

S10 An example in which noise-induced selection can never reverse the direction of evolution over short timescales, but may nevertheless affect the stationary distribution

Consider a slightly modified version of the example covered in the main text. Consider here two types in which rate modulation decreases the birth rate and increases the death rate of type 1 individuals. In equations, such modulation can be modelled via the relations:

$$b_1^{(\mathrm{ind})}(p, N_K) = b_2^{(\mathrm{ind})}(p, N_K) - \epsilon_b \tag{S97a}$$

$$d_1^{(\mathrm{ind})}(p, N_K) = d_2^{(\mathrm{ind})}(p, N_K) + \epsilon_d \tag{S97b}$$

where ϵ_b and ϵ_d are non-negative real numbers describing the effect of the ecological rate modulator on the birth and death rates respectively. Note that in this case, ϵ_b cannot be arbitrarily large: we require $\epsilon_b \leq \inf\limits_{\substack{p\in[0,1]\\ N_K\geq 0}} \{b_2^{(\mathrm{ind})}(p, N_K)\}$ to avoid negative birth rates. As in the main text, we can calculate the selection coefficient and noise-induced selection coefficient, to find

$$s(p, N_K) = -[\epsilon_b + \epsilon_d] \tag{S98a}$$

$$\kappa(p, N_K) = \epsilon_b - \epsilon_d \tag{S98b}$$

Here, s is always negative whenever there is some rate modulation in the system (*i.e.* ϵ_b and ϵ_d are not both 0), and thus natural selection always favors type 2 over type 1. Note that here, when evolution is neutral with respect to natural selection ($s = 0$), we must have $\epsilon_b = \epsilon_d = 0$. In this case, $b_1^{(\text{ind})}(p, N_K) = b_2^{(\text{ind})}(p, N_K)$ and $d_1^{(\text{ind})}(p, N_K) = d_2^{(\text{ind})}(p, N_K)$, and thus the two types are exactly equivalent in every respect.

As in the main text, we can first examine when the sign of $\mathbb{E}[dp/dt]$ is reversed relative to infinite population expectations. Since $s < 0$, we can use Eq. 14 to say the expected trajectory is in the opposite direction of infinite population predictions if $s + \kappa/KN_K > 0$. Using Eq. S98, we see that this is equivalent to

$$-[\epsilon_b + \epsilon_d] + \frac{1}{KN_K}(\epsilon_b - \epsilon_d) > 0 \Rightarrow \left(1 - \frac{1}{KN_K}\right)\epsilon_b + \left(1 + \frac{1}{KN_K}\right)\epsilon_d < 0 \tag{S99}$$

$$\Rightarrow \frac{\epsilon_b}{\epsilon_d} < -\frac{KN_K + 1}{KN_K - 1} < 0 \tag{S100}$$

Since ϵ_b and ϵ_d are both non-negative, so is their ratio, and thus inequality S100 can never be satisfied. We therefore conclude that noise-induced selection *cannot* reverse the sign of $\mathbb{E}[dp/dt]$ relative to infinite population expectations in this case.

However, noise-induced selection may still affect the long-term behaviour via the stationary distribution. We see from Eq. 15 that type 1 may be favored via noise-induced selection in the stationary distribution if dV/dp is sufficiently negative. Using the definition of V from Eq. 16b, we see that dV/dp is negative whenever

$$\frac{d\tau_1}{dp} < \epsilon_b - \epsilon_d \tag{S101}$$

Note, however, that for this system, since $E(p)$ will always be positive, $dV/dp < 0$ is a necessary but not a sufficient condition for deviation from infinite population expectations — we also require dV/dp to be large enough in magnitude relative to $E(p)$ to ensure that the RHS of Eq. 15 as a whole becomes positive.

S11 An example of non-neutral competition where evolution does not proceed in the direction of natural selection due to noise-induced effects

In this section, we provide an example of resource competition with both natural selection and mutation in which noise-induced selection reverses the direction of evolution predicted by natural selection.

Consider a community that contains two types of birds, say type 1 and type 2. These birds compete for limited resources, but in a peculiar manner: Though the two birds feed on different food sources, the trees that type 1 birds use for nesting are the same as those that the type 2 birds rely on for food. Both types are territorial and do not tolerate other individuals of either type on either their nesting or feeding sites. Thus, competition between the two types affects the *birth rate* of type 1 birds (because they can't find good nesting sites) but the *death rate* of type 2 birds (due to starvation), whereas intratype competition affects the death rate in both cases due to competition for food sources. We also assume that when individuals give birth, they may give birth to offspring of the opposite type (due to mutations) at a rate $\lambda > 0$. Thus, the influx rate λ here is a mutation rate, and we will therefore denote it by $\lambda = \mu$ to align with standard notational conventions. Let n_i be the number of type i individuals (which may vary over time). Assuming trees and birds are both randomly distributed through the landscape and the population dynamics of birds has linear density dependence, the simplest model that can incorporate these features of resource competition is given by:

$$\begin{aligned} b_1(n_1,n_2) &= n_1 - (1+\epsilon)\frac{n_1 n_2}{K} + \mu n_2 \quad &;\quad d_1(n_1,n_2) &= \frac{n_1^2}{K} \\ b_2(n_1,n_2) &= n_2 + \mu n_1 \quad &;\quad d_2(n_1,n_2) &= \frac{n_2^2}{K} + \frac{n_1 n_2}{K} \end{aligned} \tag{S102}$$

where K is a carrying capacity for the habitat, similar to Lotka-Volterra competition, and ϵ is a parameter, which as we shall see below, quantifies which type has a competitive advantage.

Moving to density space via the change of variables $x_i = n_i/K$, letting $\mathbf{x} = [x_1, x_2]^{\mathrm{T}}$, and comparing terms with Eq. 2, we see that the per-capita fitness w_i of each type is:

$$\begin{aligned} w_1(\mathbf{x}) &= 1 - x_1 - (1+\epsilon)x_2 = 1 - pN_K - (1+\epsilon)(1-p)N_K \\ w_2(\mathbf{x}) &= 1 - x_1 - x_2 = 1 - pN_K - (1-p)N_K \end{aligned}$$

where $N_K = x_1 + x_2$ is the (scaled) total population size and $p = x_1/N_K$ is the frequency of type 1 individuals in the population. In frequency space, we thus see that the selection coefficient $s := w_1 - w_2$ is given by

$$s(p, N_K) = -\epsilon(1-p)N_K \tag{S103}$$

This calculation makes it clear that ϵ controls the strength and direction of natural selection operating in the system — when $\epsilon > 0$, natural selection favors type 2, whereas when $\epsilon < 0$, type 1 is favored. When $\epsilon = 0$, the two types of birds have the same fitness and there is no natural selection operating in the system. If we now compute the per-capita turnover rates τ_i of each type, we have

$$\begin{aligned} \tau_1(\mathbf{x}) &= 1 + x_1 - (1+\epsilon)x_2 = 1 + pN_K - (1+\epsilon)(1-p)N_K \\ \tau_2(\mathbf{x}) &= 1 + x_1 + x_2 = 1 + N_K \end{aligned}$$

and the noise-induced selection coefficient $\kappa := \tau_2 - \tau_1$ is therefore

$$\kappa(p, N_K) = (2+\epsilon)(1-p)N_K \tag{S104}$$

Note that when $\epsilon = 0$, s vanishes but κ does not, meaning that the system exhibits noise-induced selection but no natural selection. Further, whenever $\epsilon > 0$ or $\epsilon < -2$, s and κ have opposite signs, *i.e.* natural selection and noise-induced selection act in opposite directions. Here, focusing on the case $\epsilon > 0$, we see from Eq. S103 that natural selection favors type 2, whereas Eq. S104 tells us that noise-induced selection favors type 1.

Finally, we also have $Q_1(\mathbf{p}) = (1-p), Q_2(\mathbf{p}) = p$. Substituting all these functional forms into Eq. 10 now tells us (after some algebra) that the frequency of type 1 individuals in the population obeys the SDE

$$\begin{aligned} dp = &\left[p(1-p)^2\left[\frac{2}{K} - \epsilon\left(N_K(t) - \frac{1}{K}\right)\right] + \mu(1-2p)\left(1 - \frac{1}{KN_K(t)}\right)\right]dt \\ &+ \frac{1}{\sqrt{KN_K(t)}}\sqrt{p(1-p)\left[1 + N_K(t)\left(1-(2+\epsilon)(1-p)^2\right)\right] + \mu\left[1-3p(1-p)\right]}dW_t \end{aligned} \tag{S105}$$

where W_t is a one-dimensional Wiener process. Upon substituting our functional forms of fitness and turnover into Eq. 12, we find that the total scaled total population size N_K obeys the SDE

$$\frac{1}{N_K}dN_K = \left[1 + \mu - N_K\left(1 + \epsilon p(1-p)\right)\right]dt + \frac{1}{\sqrt{KN_K}}\sqrt{1 + \mu + N_K\left(1 - \epsilon p(1-p)\right)}dW_t^{N_K} \tag{S106}$$

where $W_t^{N_K}$ is a one-dimensional Wiener process. We are now in a position to study the behavior of this system.

The infinite population limit

If we let $K \to \infty$, the SDE for type frequency given by Eq. S105 reduces to an ODE

$$\frac{dp}{dt} = -\epsilon N_K p(1-p)^2 + \mu(1-2p) \tag{S107}$$

When there is no mutation and no selection in the system ($\mu = \epsilon = 0$), the RHS of Eq. S107 vanishes, and thus, any initial trait frequency p_0 is expected to remain unchanged forever. If we switch off mutation alone ($\mu = 0, \epsilon \neq 0$), it is easy to check that the type favored by selection will become fixed in the population. Instead, if we switch off selection alone ($\mu \neq 0, \epsilon = 0$), mutations drive the population to a state in which both types are equally prevalent (i.e., $p = 0.5$). When both selection and mutation are present in the system, the stable fixed point in the infinite population limit will lie in $(0, 1/2)$ when $\epsilon > 0$, and will lie in $(1/2, 1)$ when $\epsilon < 0$.

Deviations from neutrality due to noise-induced selection in finite, fluctuating populations

The effects of noise-induced selection on the expected dynamics are clearest when there is no natural selection ($\epsilon = 0$) and no mutation ($\mu = 0$): In this case, the equation for trait frequencies (Eq. S105) becomes

$$dp = \frac{2p(1-p)^2}{K}dt + \frac{1}{\sqrt{KN_K(t)}}\sqrt{p(1-p)\left[1 + N_K(t)\left(1 - 2(1-p)^2\right)\right]}dW_t \tag{S108}$$

If we now take expectations on both sides, the stochastic integral term vanishes and we obtain an ODE for the expected trait frequency in the population. Assuming that derivatives and expectations commute, this ODE takes the form

$$\frac{d}{dt}\mathbb{E}[p] = \frac{2}{K}\mathbb{E}[p(1-p)^2] \tag{S109}$$

Since the RHS is always positive for $p \in (0, 1)$, we conclude that the frequency of type 1 birds is always expected to increase until type 1 becomes fixed in the population. Thus, noise-induced selection, in this case, has led to a deviation from the true neutrality in the expected dynamics — in the infinite population case, any initial trait frequency p_0 is expected to remain unchanged forever, whereas for finite, fluctuating populations, assuming $p_0 \notin \{0, 1\}$, the trait frequency of type 1 birds is expected to increase until type 1 eventually fixes in the population. Note that unlike in neutral drift, type 1 is *always* expected to be the type that becomes fixed in the population, despite the two types having the same fitness.

Reversal of the direction of evolution in finite, fluctuating populations

For the birth-death processes of the type we study in this paper, the entire population will go extinct in finite time with probability 1 (Ethier and Kurtz, 1986). Thus, the true stationary distribution for our system is thus the trivial state $x_1 = x_2 = 0$, a state at which p is undefined. However, the expected time to extinction is often so large that we can instead speak of the 'quasi-stationary distribution' of the stochastic process, obtained by only examining the system before the population goes extinct (Karlin and Taylor, 1981). Thus, we are interested in the behavior of the system in (p, N_K) space conditioned on $N_K > 0$. To study the behavior of the trait frequency when the population is far from extinction, we can simply use the naive assumption $N_K \equiv 1$ to arrive at an approximate description of the system. Under this approximation, the probability density $\pi_{qs}(p)$ of the quasi-stationary distribution is given by (see Supplementary section S6)

$$\pi_{qs}(p) = \frac{\mathcal{N}}{G(p)}\exp\left(\int_0^p \frac{F(q)}{G(q)}dq\right) \tag{S110}$$

where $\mathcal{N}$ is a normalization constant and F and G are functions given by

$$\begin{aligned} F(p) &\coloneqq p(1-p)^2\left[2-\epsilon\left(K-1\right)\right]+\mu(1-2p)\left(K-1\right) \\ G(p) &\coloneqq p(1-p)\left(2-(2+\epsilon)(1-p)^2\right)+\mu\left(1-3p(1-p)\right) \end{aligned} \tag{S111}$$

Since the above solution is an approximation, we also conduct exact stochastic individual-based simulations of the complete system defined by Eq. S102 using the Gillespie algorithm. The results of the simulations, as well as the solution predicted by Eq. S110, are plotted for a small $\epsilon > 0$ (corresponding to weak selection against type 1) in figure S1.

For low values of K and ϵ, both the stochastic individual-based simulations and the approximate solution given by Eq. S110 indicate that noise-induced selection causes the distribution of types in the population to be biased in favor of type 1 (rightmost peak in Fig S1A), a bias that disappears when K and ϵ are high (Fig S1B). To more carefully quantify when type 1 is favored by evolution, we can follow McLeod and Day (2019)’s approach and compute the quantity $\int_{1/2}^{1}\pi_{qs}(p)dp$. This quantity tells us the likelihood of observing the population in a state in which more than half of the individuals are of type 1. Since the infinite population limit from Eq. S107 predicts that $p \in (0,1/2)$ at equilibrium, we can say that the direction of evolution has been reversed relative to infinite population predictions made from natural selection and mutation alone if $\int_{1/2}^{1}\pi_{qs}(p)dp \geq 1/2$ in finite populations. The value of $\int_{1/2}^{1}\pi_{qs}(p)dp$ for various values of ϵ and K are plotted in figure S1C. For low values of ϵ (weak selection) and K (small population size), $\int_{1/2}^{1}\pi_{qs}(p)dp \geq 1/2$ and thus the direction of evolution has been reversed relative to the predictions of classical natural selection and mutation alone.

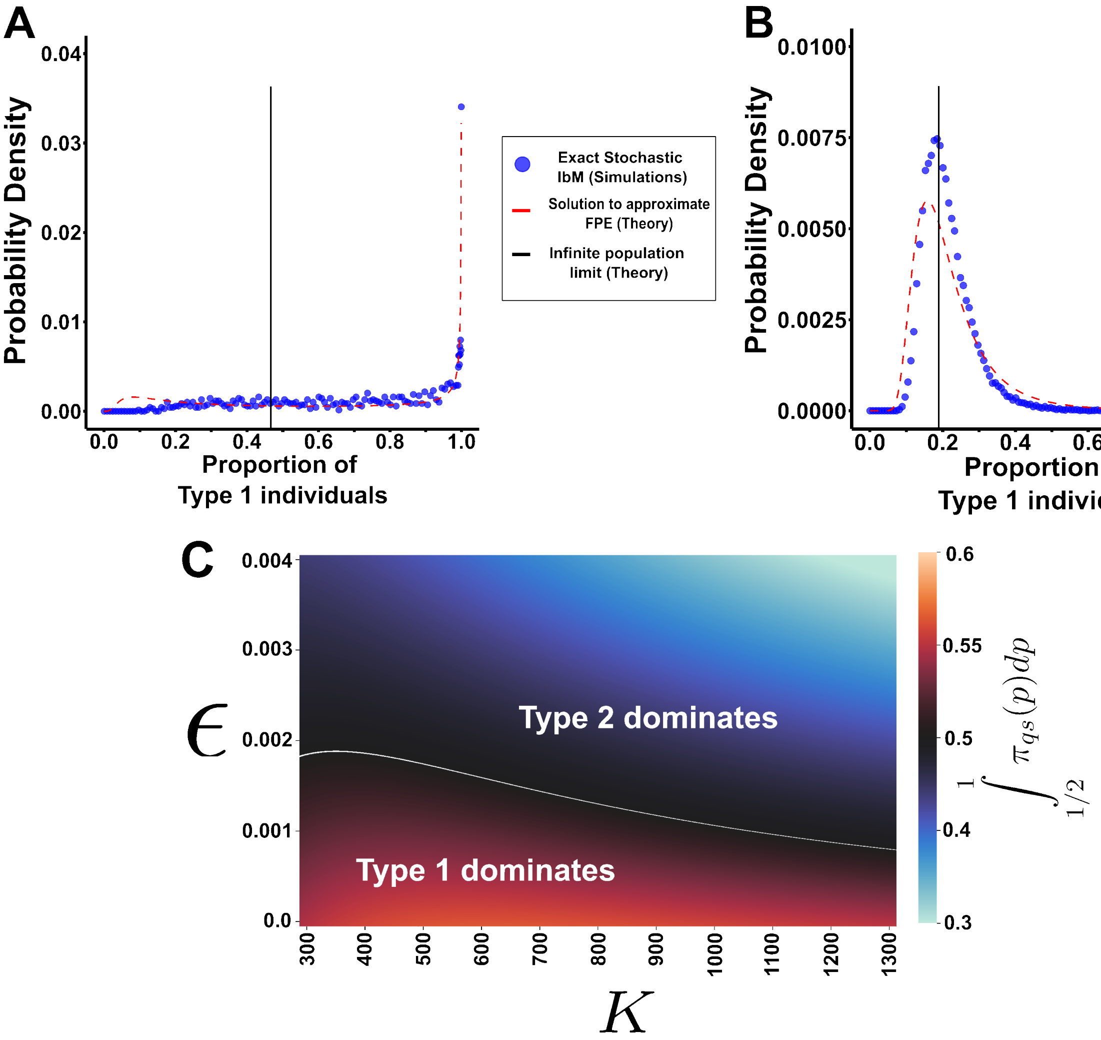

Figure S1: Predictions of our resource-competition model for various parameters. The quasi-stationary distribution has been plotted for **(A)** $K = 500, \epsilon = 0.0005$, and **(B)** $K = 5000, \epsilon = 0.005$. Blue points are from 100 independent Gillespie simulations of the exact birth-death process defined by Eq. S102, each supplied with the initial condition $n_1 = n_2 = K/2$ and allowed to run for 10^5 timesteps or until the complete population went extinct. The red dotted line is derived from numerically evaluating the RHS of equation Eq. S110. The solid black line is the infinite population limit, obtained by solving equation Eq. S107 under the approximation $N_K \equiv 1$. **(C)** A heatmap of the values of $\int_{1/2}^{1} \pi_{qs}(p)dp$ for various values of ϵ and K. If this quantity is greater than $1/2$, then type 1 is more likely to be the dominant species in the population. The white curve indicates parameter values at which $\int_{1/2}^{1} \pi_{qs}(p)dp = 0.5$, *i.e.* the population is expected to contain an equal number of type 1 and type 2 individuals. For low values of ϵ and K, type 1 is likely to be present at a higher frequency than type 2 in the population, despite deterministic natural selection predicting the opposite. For all plots in this figure, $\mu = 0.01$.

Literature Cited in Supplementary

Black, A. J. and McKane, A. J. (2012). "Stochastic formulation of ecological models and their applications". en. In: *Trends in Ecology & Evolution* 27.6, pp. 337–345. ISSN: 0169-5347. DOI: `10.1016/j.tree.2012.01.014`.

Chotibut, T. and Nelson, D. R. (2015). "Evolutionary dynamics with fluctuating population sizes and strong mutualism". In: *Physical Review E* 92.2, p. 022718. DOI: `10.1103/PhysRevE.92.022718`.

Constable, G. W. A., Rogers, T., McKane, A. J., and Tarnita, C. E. (2016). "Demographic noise can reverse the direction of deterministic selection". In: *Proceedings of the National Academy of Sciences* 113.32, E4745–E4754. DOI: `10.1073/pnas.1603693113`.

Czuppon, P. and Traulsen, A. (2021). "Understanding evolutionary and ecological dynamics using a continuum limit". en. In: *Ecology and Evolution* 11.11, pp. 5857–5873. ISSN: 2045-7758. DOI: `10.1002/ece3.7205`.

Day, T., Parsons, T. L., Lambert, A., and Gandon, S. (2020). "The Price equation and evolutionary epidemiology". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1797, p. 20190357. DOI: `10.1098/rstb.2019.0357`.

Etheridge, A. (2011). *Some mathematical models from population genetics: École d'été de Probabilités de Saint-Flour XXXIX-2009*. eng. Lecture Notes in Mathematics. Berlin: Springer-Verlag. ISBN: 978-3-642-16631-0.

Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes: characterization and convergence*. eng. Wiley series in probability and mathematical statistics. New York: Wiley. ISBN: 978-0-471-08186-9.

Frank, S. A. (2012). "Natural selection. IV. The Price equation*". en. In: *Journal of Evolutionary Biology* 25.6, pp. 1002–1019. ISSN: 1420-9101. DOI: `10.1111/j.1420-9101.2012.02498.x`.

Frank, S. A. and Slatkin, M. (1992). "Fisher's fundamental theorem of natural selection". en. In: *Trends in Ecology & Evolution* 7.3, pp. 92–95. ISSN: 0169-5347. DOI: `10.1016/0169-5347(92)90248-A`.

Hofbauer, J. and Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press. ISBN: 978-0-521-62570-8. DOI: `10.1017/CBO9781139173179`.

Houchmandzadeh, B. and Vallade, M. (2012). "Selection for altruism through random drift in variable size populations". In: *BMC Evolutionary Biology* 12.1, p. 61. ISSN: 1471-2148. DOI: `10.1186/1471-2148-12-61`.

Humplik, J., Hill, A. L., and Nowak, M. A. (2014). "Evolutionary dynamics of infectious diseases in finite populations". en. In: *Journal of Theoretical Biology* 360, pp. 149–162. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2014.06.039.

Karatzas, I. and Shreve, S. E. (1998). *Brownian motion and stochastic calculus*. eng. 2nd ed. Graduate texts in mathematics. New York: Springer-Verlag. ISBN: 978-0-387-97655-6.

Karlin, S. and Taylor, H. M. (1981). *A second course in stochastic processes*. eng. New York: Academic Press. ISBN: 978-0-12-398650-4.

Kokko, H. (2021). "The stagnation paradox: the ever-improving but (more or less) stationary population fitness". In: *Proceedings of the Royal Society B: Biological Sciences* 288.1963, p. 20212145. DOI: 10.1098/rspb.2021.2145.

Lion, S. (2018). "Theoretical Approaches in Evolutionary Ecology: Environmental Feedback as a Unifying Perspective". In: *The American Naturalist* 191.1, pp. 21–44. ISSN: 0003-0147. DOI: 10.1086/694865.

McLeod, D. V. and Day, T. (2019). "Social evolution under demographic stochasticity". en. In: *PLOS Computational Biology* 15.2, e1006739. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006739.

Øksendal, B. K. (1998). *Stochastic differential equations: an introduction with applications*. English. Berlin; New York: Springer. ISBN: 978-3-662-03620-4.

Page, K. M. and Nowak, M. A. (2002). "Unifying Evolutionary Dynamics". en. In: *Journal of Theoretical Biology* 219.1, pp. 93–98. ISSN: 0022-5193. DOI: 10.1006/jtbi.2002.3112.

Parsons, T. L., Lambert, A., Day, T., and Gandon, S. (2018). "Pathogen evolution in finite populations: slow and steady spreads the best". In: *Journal of The Royal Society Interface* 15.147, p. 20180135. DOI: 10.1098/rsif.2018.0135.

Queller, D. C. (2017). "Fundamental Theorems of Evolution". In: *The American Naturalist* 189.4, pp. 345–353. ISSN: 0003-0147. DOI: 10.1086/690937.

Van Kampen, N. G. (1981). *Stochastic processes in physics and chemistry*. eng. Amsterdam, New York, New York: North-Holland. ISBN: 978-0-444-86200-6.