Neutral theory of cooperators

Jordi Piñero,[1,*] Artemy Kolchinsky,[1,2,†] Sidney Redner,[3,‡] and Ricard Solé[1,3,4,§]

[1]*ICREA-Complex Systems Lab, Universitat Pompeu Fabra, 08003 Barcelona, Spain*
[2]*Universal Biology Institute, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[3]*Santa Fe Institute, Santa Fe, New Mexico 87501, United States*
[4]*Institut de Biologia Evolutiva (CSIC-UPF), 08003 Barcelona, Spain*

Mutualistic interactions are widespread in nature, from plant communities and microbiomes to human organizations. Along with competition for resources, cooperative interactions shape biodiversity and contribute to the robustness of complex ecosystems. We present a stochastic neutral theory of cooperator species. Our model shares with the classic neutral theory of biodiversity the assumption that all species are equivalent, but crucially differs in requiring cooperation between species for replication. With low migration, our model displays a bimodal species-abundance distribution, with a high-abundance mode associated with a core of cooperating species. This core is responsible for maintaining a diverse pool of long-lived species, which are present even at very small migration rates. We derive analytical expressions of the steady-state species abundance distribution, as well as scaling laws for diversity, number of species, and residence times. With high migration, our model recovers the results of classic neutral theory. We briefly discuss implications of our analysis for research on the microbiome, synthetic biology, and the origin of life.

1. INTRODUCTION

Cooperation is ubiquitous across scales in complex systems. In ecology, cooperative interactions shape ecosystem structures [1, 2]. Within biology, mutualism (a reciprocal form of cooperation) is the engine of evolutionary transitions [3, 4] and constitutes an essential part of the architecture of biodiversity [5, 6]. In microbiomes, cooperative interactions occur through extensive cross-feeding exchanges associated with shared diffusive metabolites [7–9], a phenomenon dubbed the *social network* of microorganisms [10]. Moreover, cooperative interactions enhance community stability [11] and facilitate metabolic functions [12]. Cooperation has also been studied in human organizations, for example, between companies engaged in jointly manufacturing a certain product [13].

Ecological theory has traditionally studied either *(i)* the dynamics of large randomly-assembled communities, leading to general stability-complexity principles [14, 15], or *(ii)* the stationary properties of stochastic interactions between neutral species, inspired by the neutral theory of biodiversity [16, 17]. Stability patterns in mutualistic communities under approach *(i)* have been addressed using network models [18–20]. The results of these models may depend strongly on the assumptions made about different species (in terms of size, lifestyle, or physiology) and their interactions, making it challenging to derive general theoretical lessons.

In contrast, approach *(ii)* assumes that all species in the system are equivalent, shifting the focus to the effects of demographic noise as the main determinant of emerging ecological patterns [21–25]. Despite its simplicity, neutral theory successfully accounts for many relevant statistical patterns, including abundance distributions, species-area relations, and diversity estimates in space and time [16, 22, 26, 27]. However, a neutral theory that explicitly involves cooperative interactions is lacking, and little is known about the impact of mutualistic exchanges under a neutral picture. For example, microbiomes exhibit marked quantitative patterns that diverge significantly from those predicted by classic neutral theory, such as the presence of a persistent subset of species [12, 28, 29] or bimodal abundance distributions [30].

In this paper, we present a neutral theory for ecosystems of cooperators. Our model preserves the neutral hypothesis — that all species in the system follow the same replication rules, albeit possibly with frequency-dependent replication probabilities. However, it departs from the classic neutral theory of biodiversity by requiring cooperation between species for replication to occur. Our model contains only two independent parameters: the total number of individuals in the system N, typically assumed to be large, and the migration rate μ, the probability that a new species enters the system at a given time step. One of the remarkable features of our model is the emergence of a bimodal species-abundance distribution, characterized by a "core" of cooperator species that remain at high abundances for much longer compared to species in the low-abundance mode. This core leads to the maintenance of species diversity even at very small migration rates, where, in the absence of cooperation, the system would rapidly fixate. In this sense, we show that cooperative interactions can provide a powerful stabilizing effect.

We solve our model analytically and confirm the results via extensive numerical simulations. Our analysis is split into two main parts: first, we focus on the statistical patterns displayed by the system's steady state; afterward, we consider the stochastic dynamics of abundance trajectories as species enter and leave the system.

2. MODEL SETUP

Our model considers a well-mixed population of N individuals, each labeled by a species. The configuration of the

* jpinerfe@gmail.com; Equal contribution.
† artemyk@gmail.com; Equal contribution.
‡ redner@santafe.edu
§ ricard.sole@upf.edu; Corresponding author.

system at a given time is represented by the *abundance vector* $\mathbf{n} = (n_1, n_2, \dots)$, where n_i is the abundance of species i. We use $R(\mathbf{n})$ to indicate the number of nonzero elements of vector $\mathbf{n}$, i.e., the number of present species. The configuration evolves stochastically over a sequence of discrete steps involving either a cooperative replication or a migration event:

1) *Replication with probability* $1-\mu$, *Fig. 1 (left)*: two individuals are randomly chosen from the population, individual A from species i and individual B from species j. If A and B belong to the same species ($i = j$), nothing happens. Otherwise, an individual C of species k is randomly selected from the population and replaced by a new individual of species i. The abundance vector $\mathbf{n}$ is updated as

$$(n_i, n_k) \to (n_i + 1, n_k - 1)\,. \tag{1}$$

(If $i = k$, there is no change). Given our well-mixed population, the probability that this event occurs in a given step is

$$(1-\mu)\frac{n_i}{N}\frac{N-n_i}{N}\frac{n_k}{N}. \tag{2}$$

In (2), the factors represent the probability that: replication takes place, A belongs to species i, B belongs to any species other than i, and C belongs to species k.

2) *Migration with probability* μ, *Fig. 1 (right)*: an individual A of species i is randomly chosen and replaced by an individual of a new species. The abundance vector $\mathbf{n}$ is updated as

$$(n_i, n_{R(\mathbf{n})+1}) \to (n_i - 1, 1)\,. \tag{3}$$

This event happens with probability $\mu(n_i/N)$ per step. For notational convenience, if any species becomes extinct ($n_i \to 0$), the species are reindexed so that the first $R(\mathbf{n})$ entries of $\mathbf{n}$ are strictly positive.

Total population size is conserved under the above rules ($\sum_i n_i = N$), resulting in competition between species for limited space. Also, because every migration event introduces a new species, migration represents an inflow from an infinitely-diverse external reservoir.

In the following, we consider the stationary properties of this stochastic process. We are particularly interested in the *species abundance distribution* P_n (the probability that a species has abundance n in steady state) and the system's overall diversity. We measure diversity using the Simpson index [31], defined for abundance vector $\mathbf{n}$ as[32]

$$\lambda(\mathbf{n}) := \sum_{i=1} \left(\frac{n_i}{N}\right)^2 . \tag{4}$$

The Simpson index is the probability that two randomly chosen individuals belong to the same species [31], and it is commonly used in ecology as an inverse measure of diversity [33]. It is bounded as $1/R(\mathbf{n}) \le \lambda(\mathbf{n}) \le 1$, with the lower bound achieved when the population is evenly distributed among $R(\mathbf{n})$ species (maximum diversity) and the upper bound achieved when the population is fixated on a single species (minimum diversity). The *inverse Simpson index* $1/\lambda(\mathbf{n})$ may be interpreted as the "effective" number of species in the system.

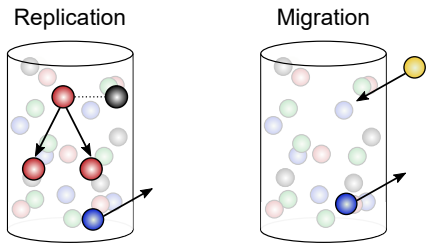

Figure 1. **Neutral model of cooperators** (species indicated by colors). The population evolves according to two rules: **(a)** during replication, individuals of different species (black and red balls) cooperate. The red individual replicates and replaces a randomly chosen individual (blue ball). **(b)** during migration with probability μ, an individual from a new species (yellow ball) enters from an external pool and substitutes a randomly-chosen individual (blue ball).

As shown below, our model has two different steady-state regimes, depending on the migration rate. When migration is high, diversity is very large ($\lambda \approx 0$), competition for space is strong, and the species abundance is well-described by the Fisher's Logseries distribution. This regime reduces to the classic neutral model of Hubbell, which does not involve cooperative interactions [34]. When migration is low, the system forms a "core" of high-abundance species, and the species-abundance distribution acquires a characteristic bimodal shape, shown schematically in Fig. 2. In what follows, we will study the emergence and nature of this core.

3. STEADY-STATE DISTRIBUTION

To derive the steady-state species abundance distribution, we exploit the neutrality of our model, which allows us to treat all present species as statistically equivalent. Thus, without loss of generality, we consider species $i = 1$ as the *representative species*.

The stochastic dynamics of the abundance of the representative species, written $n \equiv n_1$, follows a birth-death process. The birth probability ($n \to n+1$ for $n \in \{1, \dots, N-1\}$) during a single time step is

$$b_n = (1-\mu)\frac{n}{N}\left(1 - \frac{n}{N}\right)^2 , \tag{5}$$

while the death probability ($n \to n-1$ for $n \in \{1, \dots, N\}$) is

$$d_n = (1-\mu)\frac{n}{N}\left(1-\frac{n}{N}\right)\left[1 - \left(1-\frac{n}{N}\right)\lambda_\circ(\mathbf{n})\right] + \mu\frac{n}{N}\,. \tag{6}$$

See SM1 for a detailed derivation of Eqs. (5)-(6). The quantity $\lambda_\circ(\mathbf{n})$ refers to the Simpson index of the other species (i.e., excluding the representative species $i = 1$):

$$\lambda_\circ(\mathbf{n}) := \sum_{i=2} \left(\frac{n_i}{N-n}\right)^2 . \tag{7}$$

Although different species interact during replication, the stochastic dynamics of each species depends on all others only through a single number, $\lambda_\circ(\mathbf{n})$. We remark that similar coupled birth-death processes have been considered in work on interacting particle systems and nonlinear chemical master equations [35].

The above birth-death process is absorbing into the extinction state $n = 0$, since any particular species will eventually go extinct. To study the abundance of non-extinct species, we may make the process ergodic by adding a positive birth probability $b_0 > 0$ out of the extinction state $n = 0$. In the following, we focus on the steady-state distribution restricted to positive abundances $n \in \{1, 2, \dots\}$, which does not depend on the choice of b_0.

We now find the steady-state species abundance distribution P_n of the representative species. Since births and deaths must balance in steady state, $P_n b_n = P_{n+1} d_{n+1}$ for $n \in \{1, \dots, N-1\}$, we have

$$P_n \propto \prod_{k=1}^{n-1} \frac{b_k}{d_{k+1}} . \tag{8}$$

This is not yet a closed equation because the death probability (6) depends on the fluctuating quantity $\lambda_\circ(\mathbf{n})$, i.e., the Simpson index of the non-representative species. However, when the number of species is large, $\lambda_\circ(\mathbf{n}) \approx \lambda(\mathbf{n})$. Futhermore, due to self-averaging, $\lambda(\mathbf{n})$ is tightly peaked around its expected steady-state value,

$$\lambda_\circ(\mathbf{n}) \approx \lambda^* := \langle \lambda(\mathbf{n}) \rangle, \tag{9}$$

where $\langle \cdot \rangle$ indicates steady-state expectation over the entire population. Plugging (9) into (6) and using (8) we find that, under reasonable assumptions, the steady-state abundance distribution is approximated as (see SM2 for details):

$$P_n \propto \frac{(1-\mu)^n}{n} e^{-(n-N\lambda^*)^2/2N} . \tag{10}$$

(10) is one of our main results, showing how the abundance distribution depends on the population size N, migration rate μ, and the steady-state Simpson index λ^*. In reality, λ^* itself depends on N and μ — the only two parameters that describe the system — although finding the explicit expression of λ^* in terms of N and μ is not trivial (we will do so in the next section). For now, we note that the Simpson index λ^* and the migration probability μ move in opposite directions. When μ decreases to 0 (migration vanishes), λ^* increases to 1 (fixation is reached). Conversely, when μ increases to 1 (only migration occurs), λ^* decreases to its minimum value of $1/N$ (maximum diversity).

Observe that (10) expresses P_n as the product of two distributions, as illustrated in Fig. 2. The first is the Logseries distribution, $(1-\mu)^n/n$, the abundance distribution of Hubbell's neutral model, which does not have cooperative interactions [36]. This contribution represents migration-driven competition for space, and it dominates the system at high migration rates. The second is a Gaussian distribution, $e^{-(n-N\lambda^*)^2/2N}$, with mean abundance $N\lambda^*$ and standard deviation $\sqrt{N}$. This contribution represents a high-abundance mode that emerges due to cooperative interactions. We term this contribution the *cooperator core*.

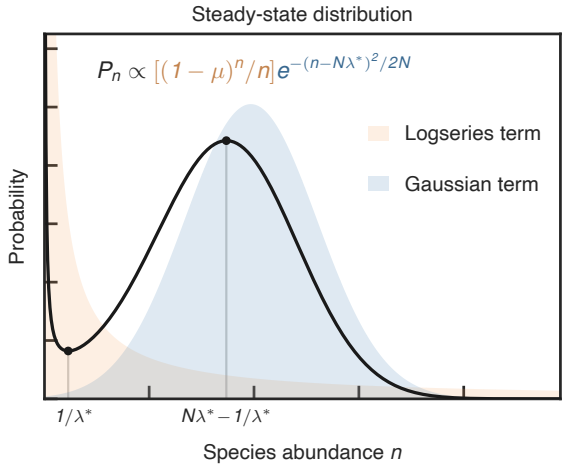

Figure 2. **Schematic illustration of the species abundance distribution** P_n in steady state. This distribution depends on population size N, migration rate μ, and the steady-state Simpson index λ^*. In Eq. (10), it is approximated as the product of Logseries distribution, arising from competition for space, and a Gaussian distribution, arising from cooperative interactions. At low migration, the combined distribution exhibits a bimodal shape, with local maxima at $n = 1, n = N\lambda^* - 1/\lambda^*$ and a local minimum at $n = 1/\lambda^*$, see Eq. (15). Note that the two terms combine multiplicatively, not additively, so their areas do not add up to the area of the combined distribution.

The emergence of the cooperator core is a highly nontrivial phenomenon, and it implies various other interesting aspects of this model (maintenance of high diversity at low migration, long residence times, etc.). In simple terms, the cooperator core arises due to the frequency-dependent (i.e., abundance-dependent) replication rates. Individuals from low-abundance species are unlikely to be randomly paired with an individual from the same species, thus they have a high per-capita probability of replicating. Conversely, individuals from high-abundance species are more likely to be randomly paired with an individual from the same species, thus they have a lower per-capita replication probability. As a result, species feel an effective force towards an intermediate abundance at $n \approx N\lambda^*$, leading to a bimodal abundance distribution as in Fig. 2.

Importantly, the cooperative core only emerges at low migration rates. At high migration, when μ is large and λ^* is low, most species exhibit low abundances near $n \approx 1$. In this regime, there is very little variation in species abundances, so the frequency dependence of the replication rates is negligible, and our system approaches Hubbell's neutral model. We also remark that bimodality requires that the mean of the Gaussian contribution ($N\lambda^*$) be many standard deviations ($\sqrt{N}$) away from the origin or, in other words, that $\lambda^* \gg 1/\sqrt{N}$. Clearly, this cannot occur at very high migration probability, since λ^* decreases to its minimum value of $1/N$ as μ increases to 1. In the next section, we will derive the explicit migration probability μ that allows for bimodality.

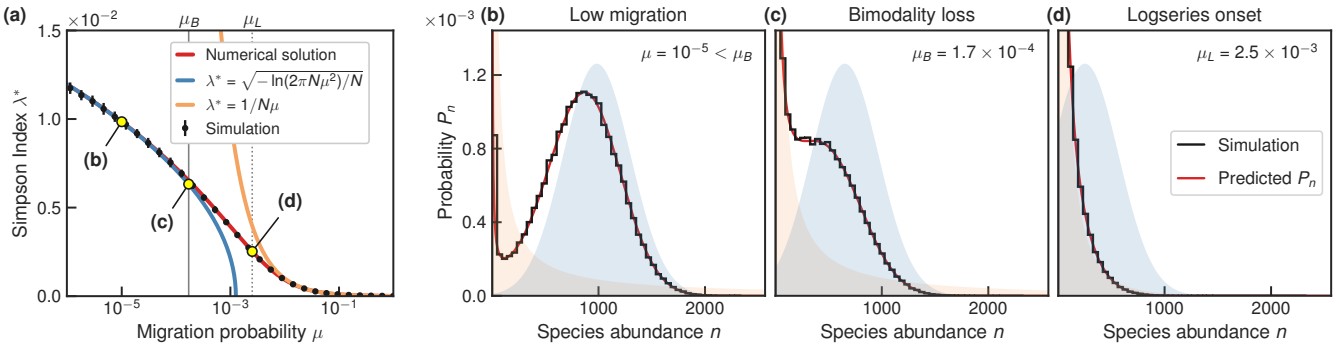

Figure 3. **Scaling of Simpson index and species abundance distributions.** **(a)** Scaling of Simpson index versus migration probability μ in a system with $N = 10^5$ individuals. We compare mean and standard deviations from simulations of the full system (100 runs; black), numerical inversion of Eq. (12) (red), approximations in low migration (13) (blue) and high migration (18) (orange) regimes. Solid vertical line indicates μ_B (16) where bimodality is lost; dotted vertical line indicates μ_L (17) where system transitions to Logseries regime. **(b)-(d)** Empirical histograms from simulations (across 1000 runs) versus predicted steady-state distributions (10) for three migration probabilities. As in Fig. 2, shaded areas represent Logseries (orange) and Gaussian (blue) contributions. **(b)** Low-migration regime exhibits a bimodal distribution with a cooperator core. **(c)** Bimodality disappears once migration probability increases beyond μ_B (16). **(d)** At higher migration probabilities, including the Logseries transition point where $\mu_L = \lambda^*$ (17), the distribution approaches the Logseries, as predicted by Hubbell's neutral theory.

In this section, we derive several important properties of the steady-state species abundance distribution and the cooperator core. In particular, we derive expressions of the steady-state Simpson index, the characteristic migration probability μ_B below which the formation of the cooperator core takes place, and the expected number of species in steady state. Most of our analyses will consider separately the case of low migration (when the core is present) and high migration. Fig. 3 illustrates our results, including the scaling of λ^* with the migration probability μ and a plot of three steady-state distributions.

A. Steady-state Simpson index

In this section, we find the Simpson index as a function of μ and N. We begin by writing λ^* in terms of the species abundance distribution P_n as

$$\lambda^* \approx \frac{1}{N}\frac{\sum_{n=1}^{N} P_n n^2}{\sum_{n=1}^{N} P_n n}. \tag{11}$$

This expression, derived in SM3, implies that the expected Simpson index is proportional to the ratio of the second and first moments of P_n. Since P_n itself depends on λ^*, we must solve (11) using self-consistency. We do this separately in the low-migration and the high-migration regimes. We also consider several important transition points.

1. *Low migration*

As shown in the SM3, in the low-migration regime, we may approximate the two sums in (11) by integrals. This leads to the equation

$$\frac{1}{\mu} = \sqrt{\frac{\pi N}{2}}\,\mathrm{erfcx}\left(\sqrt{\frac{N}{2}}(\mu - \lambda^*)\right), \tag{12}$$

where $\mathrm{erfcx}(z) := e^{z^2}\mathrm{erfc}(z) = e^{z^2}\sqrt{4/\pi}\int_z^\infty e^{-t^2}\mathrm{d}t$ is the *scaled complementary error function* [37]. Although the function $\mathrm{erfcx}(z)$ does not have a closed-form inverse, it has an efficient numerical implementation [38] that allows us to quickly compute λ^* using numerical root-finding algorithms. The scaling predicted by solving this equation for fixed N and varying μ is shown as a red curve in Fig. 3(a).

A closed-form approximation is possible for sufficiently small μ. For $\mu - \lambda^* \ll 0$, we may use $\mathrm{erfcx}(z) \approx 2e^{z^2}$ for $z \to -\infty$. Plugging into (12) and solving gives

$$\lambda^* \approx \mu + \sqrt{\frac{-\ln(2\pi N\mu^2)}{N}} \approx \sqrt{\frac{-\ln(2\pi N\mu^2)}{N}}. \tag{13}$$

Note that this approximation is only defined for $\mu \le 1/\sqrt{2\pi N}$, so that the argument of the logarithm is nonnegative.

(13) implies that, in the low-migration regime, the Simpson index scales as $\sim 1/\sqrt{N}$ in population size and as $\sim \sqrt{-\ln\mu}$ in the migration probability. The scaling predicted by (13) is shown as a blue curve in Fig. 3(a). The predicted abundance distribution is compared against simulations in Fig. 3(b).

Interestingly, an expression similar to Eq. (12) has previously appeared in a different ecological model, which considers a population embedded in a fluctuating spatiotemporal fitness landscape [39, Appendix C].

2. *Bimodality point*

We now find the migration probability below which the species abundance distribution becomes bimodal. To do so,

we consider P_n (10) as a differentiable function of n and find the critical abundances $\hat{n}$ where its derivative vanishes. With some algebra, these are found to be

$$\hat{n} = \frac{N}{2}\left[\lambda^* + \ln(1-\mu) \pm \sqrt{(\lambda^* + \ln(1-\mu))^2 - 4/N}\right]. \tag{14}$$

We simplify by expanding around large N and small μ to express the two critical abundances as

$$\hat{n}_{\min} \approx \frac{1}{\lambda^*} \quad \text{and} \quad \hat{n}_{\max} \approx N\lambda^* - \frac{1}{\lambda^*}, \tag{15}$$

shown as the local minimum and local maximum in Fig. 2. These two points become distinct when

$$\lambda^* \geq \frac{2}{\sqrt{N}} \quad \text{and} \quad \mu \leq \mu_B := \frac{e^{-2}}{\sqrt{2\pi N}}, \tag{16}$$

where μ_B is found by plugging $\lambda^* = 2/\sqrt{N}$ into (13) and solving. Thus, μ_B is the migration probability below which we see the formation of a bimodal abundance distribution.

3. *Logseries onset*

Another interesting value for the migration probability separates the cooperation- from the migration-dominated regimes. As we increase μ towards 1, the abundance distribution becomes dominated by the Logseries term, recovering the distribution predicted by Hubbell's neutral theory of biodiversity.

To identify this transition point, we consider the migration rate at which μ becomes larger than λ^*. Plugging $\mu - \lambda^* = 0$ into (12) and using $\mathrm{erfcx}(0) = 1$ specifies this point as

$$\mu_L := \sqrt{\frac{2}{N\pi}}. \tag{17}$$

4. *High migration*

As mentioned above, when μ is large, the abundance distribution approaches that of Hubbell's neutral model. To estimate the steady-state Simpson index λ^* in this regime, observe that the term $(1-\mu)^n$ in (10) imposes an exponential cutoff on abundances $n > 1/\mu$. Considering the regime $\mu > \mu_L$ from (17), we may restrict our attention to $n < 1/\mu < \sqrt{N\pi/2}$ and $\lambda^* < \mu$. For these values, the Gaussian term can be approximated as a constant factor, $e^{-(n-N\lambda^*)^2/2N} \approx e^{-N\lambda^{*2}/2}$. We re-evaluate λ^* (11) using the Logseries distribution $P_n \propto (1-\mu)^n/n$ and large N, giving

$$\lambda^* \approx \frac{1}{N\mu}. \tag{18}$$

Thus, with high migration, the Simpson index scales as $\lambda^* \sim N^{-1}$ in population size and $\lambda^* \sim \mu^{-1}$ in migration probability. This scaling is shown with the orange curve in Fig. 3(a).

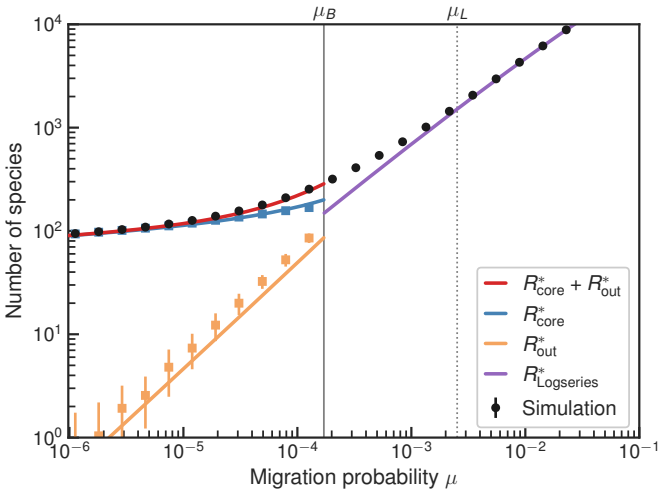

Figure 4. **Scaling of number of species R^* versus migration probability μ** ($N = 10^5$). We compare mean and standard deviations from simulations of the full system (100 runs; black); approximations in the low-migration regime of the number of total species (red), core species (24) (blue), and non-core species (25) (orange); approximations in the high-migration regime of the total number of species (purple). Solid vertical line indicates μ_B (16) where bimodality is lost; dotted vertical line indicates μ_L (17) where system transitions to Logseries regime.

In fact, as shown in SM3 B, the high-migration expression (18) can be derived from (12). This is surprising because (12) was derived under the assumption of $\mu \ll \lambda$. Empirically, we found that the value of λ^* given by the numerical solution to (12) provides an excellent match to λ^* calculated from simulations, even for large migration probabilities (see Fig. 3(a), red curve).

B. Steady-state number of species

As another important measure of diversity, we consider the expected number of different species in steady state:

$$R^* := \langle R(\mathbf{n}) \rangle. \tag{19}$$

The expected steady-state abundance of the representative species obeys $\sum_n P_n n = N/R^*$, therefore the number of species can be found as

$$R^* = \frac{N}{\sum_n P_n n}. \tag{20}$$

With high migration, R^* can be estimated by evaluating the denominator of (20). We may ignore the Gaussian contribution in (10), since the most relevant contribution to (10) is the Logseries distribution. For large N, the normalization constant of the Logseries term is $-\ln\mu$, thus we compute

$$\sum_n P_n n \approx \frac{1}{-\ln\mu}\sum_n (1-\mu)^n \approx \frac{1-\mu}{-\mu\ln\mu}. \tag{21}$$

Substituting back into (20) gives

$$R^* \approx R^*_{\text{Logseries}} := -\frac{\mu N}{1-\mu} \ln \mu, \tag{22}$$

shown in purple in Fig. 4. As expected, in the limit of $\mu \to 1$, $R^*_{\text{Logseries}} \to N$. This indicates that if only migration occurs (with no replication taking place), then the system acquires the maximum number of species with each of the N individuals belonging to a different species.

In the low-migration regime, below the point of bimodality (16), there are two relevant species counts. The first is R^*_{core}, the number of high-abundance species that belong to the cooperator core. We define these species as those with abundances larger than the local minimum, $n > 1/\lambda^*$, see Fig. 2. The second is R^*_{out}, the number of low-abundance species that remain outside of the cooperator core and have abundance $n \leq 1/\lambda^*$. The total number of species is given by

$$R^* \approx R^*_{\text{core}} + R^*_{\text{out}}\,. \tag{23}$$

Let us introduce N_{core} and N_{out} as the total number of individuals inside and outside the core, respectively, where $N_{\text{core}} + N_{\text{out}} = N$. We also denote by $\langle n \rangle_{\text{core}}$ and $\langle n \rangle_{\text{out}}$ the expected abundances of a species conditioned to be inside and outside the core, respectively. To estimate R^*_{core}, we use the relation $R^*_{\text{core}} = N_{\text{core}}/\langle n \rangle_{\text{core}}$. As shown in SM4, to a first approximation we may take $N_{\text{core}} \approx N$ (i.e., in the low-migration regime, most individuals belong to the core). Then, using $\langle n \rangle_{\text{core}} \approx \hat{n}_{\text{max}}$ from (15), we obtain

$$R^*_{\text{core}} \approx \frac{1}{\lambda^* - (N\lambda^*)^{-1}}, \tag{24}$$

shown in blue in Fig. 4. Similarly, we estimate R^*_{out} by using $R^*_{\text{out}} = N_{\text{out}}/\langle n \rangle_{\text{out}}$. We may estimate this ratio as (see SM4):

$$R^*_{\text{out}} \approx -N\mu \ln \lambda^*, \tag{25}$$

shown in orange in Fig. 4. The red curve shows the total number of species $R^* \approx R^*_{\text{core}} + R^*_{\text{out}}$. Observe that at low migration μ, almost all species belong to the core.

4. DYNAMICAL PROPERTIES

We now study the stochastic dynamics of our model by considering trajectories of species as they enter and leave the population. We focus mainly on the low-migration regime, where cooperation plays an important role.

The dynamics of any given species is described by a trajectory of nonnegative abundance values $n(t)$, starting from $n(t_0) = 1$ when that species enters the population at time $t = t_0$ and ending on $n(t_f) = 0$ when that species goes extinct at time $t = t_f$. We characterize each trajectory by two statistics: the maximum abundance reached, $\max_{t:t_0 \leq t \leq t_f} n(t)$, and the residence time before extinction, $t_f - t_0$.

For concreteness, we illustrate two typical abundance trajectories in Fig. 5. We see that the blue trajectory resides in the system for a very long time and reaches a high maximum abundance value, while the orange trajectory does not reach a high abundance and quickly goes extinct.

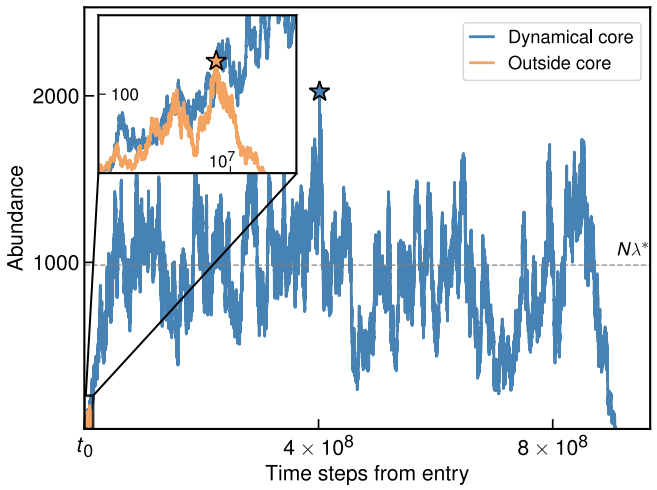

Figure 5. **Abundance trajectories from entry to extinction**, illustrating that species that enter the dynamical core reside for much longer times. We sample two abundance trajectories from the stationary process ($N = 10^5, \mu = 10^{-5}$), one for a species that enters the dynamical core and one for a species that does not. Dashed gray line indicates the abundance position of the local minimum discussed in Sec. 4 4 A, stars indicate maximum abundances reached.

In fact, simulations show that all trajectories cluster into two well-defined classes: one with long residence times and high abundances, which we term the *dynamical core*, and one with short residence times and low abundances. To illustrate this, Fig. 6, shows a scatter plot of the maximum abundances and residence times of 10^6 randomly sampled species. Species that fall into the dynamical core (top right cluster) tend to have residence times orders of magnitude larger than those that do not enter the dynamical core. Moreover, the distribution of maximum abundances has a clear bimodal shape (inset), allowing us to define a quantitative threshold for classifying species as belonging to the dynamical core.

In the following, we derive the distribution of maximum abundances, shown in red in Fig. 6. We then use this distribution to define the abundance threshold for the dynamical core. We also derive the mean residence times of species that belong and do not belong to the dynamical core.

A. Maximum abundances

Let Q_m indicate the probability that a typical species trajectory reaches a maximum abundance m. To find the form of Q_n, it will be helpful to introduce the quantity $u_n(\ell)$ to represent the probability that a species with abundance n reaches abundance ℓ or higher before going extinct. Q_m and $u_n(\ell)$ are related by

$$Q_m = u_1(m) - u_1(m+1)\,. \tag{26}$$

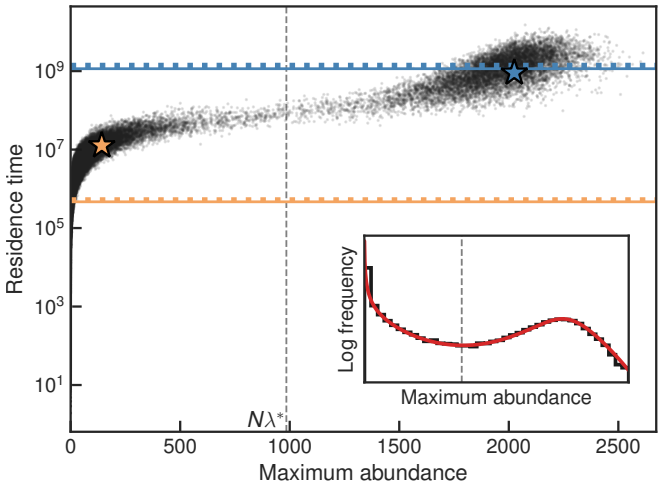

Figure 6. **Species residence times and maximum abundances** for 10^6 species sampled from stationary dynamics ($N = 10^5$, $\mu = 10^{-5}$). The species fall into two clusters: one with long residence times and high maximum abundances, the *dynamical core*, and one with short residence times and low maximum abundances. The residence times and maximum abundances of the two trajectories from Fig. 5 are indicated with star symbols. Horizontal lines indicate predicted (solid) and empirical (dashed) mean residence times of species in the dynamical core (blue) and outside of the dynamical core (orange). Inset: the empirical histogram (black) and prediction (red, from Eqs. (26) and (28)) of maximum abundances distribution Q_m. The distribution has a bimodal shape, with the dividing point located near abundance $N\lambda^*$ (dashed vertical line in main plot and inset).

In other words, Q_m is equal to the probability that a species that enters the system at abundance $n = 1$ eventually reaches abundance m but not abundance $m + 1$.

Importantly, the birth-death process defined by (5)-(6), $u_n(\ell)$ obeys the following recurrence:

$$u_n(\ell) = b_n u_{n+1}(\ell) + d_n u_{n-1}(\ell) + (1 - b_n - d_n)u_n(\ell). \tag{27}$$

The boundary conditions are $u_0(\ell) = 0$ (extinct species never come back into the system) and $u_\ell(\ell) = 1$. A general solution to the recurrence (27) is derived in SM5. For $u_1(\ell)$, the quantity that enters into (26), this solution is approximated as

$$u_1(\ell) \approx \left[e^{\ell^2/2N-(\ell-1)\lambda^*}\left(\sqrt{2N}\mathscr{D}(x_\ell) + \tfrac{1}{2}\right) - \sqrt{2N}\mathscr{D}(x_1) + \tfrac{1}{2}\right]^{-1}, \tag{28}$$

where we introduce the rescaled abundances $x_\ell := (\ell - N\lambda^*)/\sqrt{2N}$ and the Dawson function $\mathscr{D}(z) := e^{-z^2}\int_0^z e^{t^2}dt$.

Plugging this result into (26) gives the probability Q_m that a species reaches maximum abundance m. The predicted and empirical distribution of maximum abundances is shown in Fig. 6 (inset). In addition, the distribution Q_m is bimodal, with a local minimum located at approximately $N\lambda^*$ (see SM5). This bimodality allows us to classify species into two sets: those whose abundances reach $N\lambda^*$ (or higher) and those that do not. We refer to the former as the *dynamical core*. For example, Fig. 5 shows a blue trajectory that reaches abundance values larger than $N\lambda^*$ (dashed line) and thus belongs to the dynamical core, while the orange trajectory does not reach this threshold and thus remains outside the dynamical core.

Finally, we use the term *infiltration probability* β to refer to the probability that a new species will enter the dynamical core. Interestingly, as we show in SM5, this probability is approximately equal to the Simpson index,

$$\beta \approx \lambda^*, \tag{29}$$

suggesting that new species are less likely to enter the dynamical core in more diverse populations.

B. Residence times

We now derive the mean residence times by separately considering species that enter the dynamical core (τ_{core}), that do not enter the dynamical core (τ_{out}), and all species (τ). For simplicity, we will also use our calculations of the expected number of all present species R^* (22)-(23), core species R^*_{core} (24), and non-core species R^*_{out} (25).

Here, there is a subtle point to be raised. Our calculations of R^*_{core} and R^*_{out} were based on a "static" definition of the core, as the set of species that have abundance greater or smaller than $1/\lambda^*$ at a single point in time. This is different from the "dynamic" definition of the core considered in this section, as the set of species whose trajectory reaches an abundance greater than $N\lambda^*$ at any point in time. In SM6, we show that mean residence times can be alternatively derived using a mean first-passage time (MFPT) calculation that does not explicitly invoke Eqs. (22)-(25). This alternative MFPT-based analysis, which we do not include here for simplicity, leads to the same quantitative results. This suggests that the two definitions of the core are essentially equivalent when restricted to the set of species present in steady state.

We now consider the low-migration regime, and derive the mean residence times of species that enter the core, τ_{core}, and those that do not enter the core, τ_{out}. The rate at which new species migrate into the system and eventually enter the dynamical core is given by $\mu\beta$. On the other hand, the rate at which core species go extinct is $R^*_{\mathsf{core}}/\tau_{\mathsf{core}}$, where R^*_{core} is the expected number of species in the core at any one time. Since these rates must balance in steady state, we have $\mu\beta = R^*_{\mathsf{core}}/\tau_{\mathsf{core}}$. Using Eqs. (24) and (29), we arrive at

$$\tau_{\mathsf{core}} \approx \frac{1}{\mu\lambda^*\left[\lambda^* - (N\lambda^*)^{-1}\right]}. \tag{30}$$

Similarly, for species that never reach the core, entry rates must be balanced against extinction rates as $\mu(1-\beta) = R^*_{\mathsf{out}}/\tau_{\mathsf{out}}$. Using Eqs. (25) and (29), this gives

$$\tau_{\mathsf{out}} \approx -\frac{N}{1-\lambda^*}\ln\lambda^* \tag{31}$$

We estimate the overall mean residence time by combining these results with the infiltration probability as

$$\tau \approx \beta\tau_{\mathsf{core}} + (1-\beta)\tau_{\mathsf{out}}. \tag{32}$$

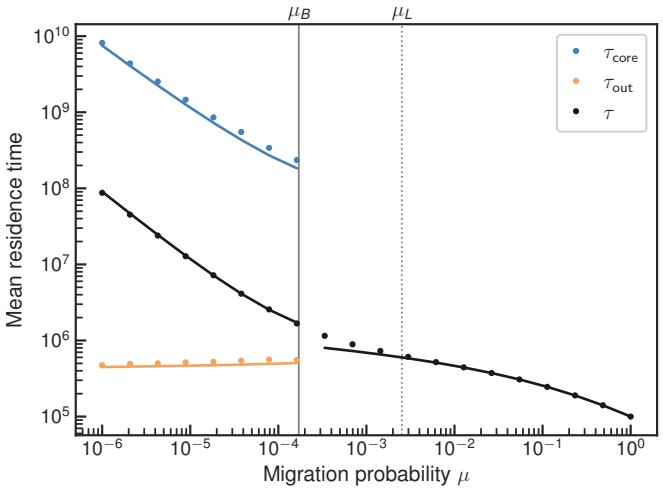

Figure 7. **Scaling of mean residence times** with migration probability μ ($N = 10^5$). Simulations (dots) are compared to predictions (solid curves) for: core species, τ_{core} (30); non-core species, τ_{out} (31); and all species, τ (32)-(33) for low and high values of μ. Solid vertical line indicates μ_B (16) where bimodality is lost; dotted vertical line indicates μ_L (17) where system transitions to Logseries regime.

These predictions are compared against data in Fig. 6 (solid and dotted horizontal lines).

The scaling of mean residence times against migration probability μ is shown in Fig. 7. For low migration probabilities, core species live orders of magnitude longer than species that do not enter the core.

Our analysis above has mostly focused on the low-migration regime, in which the distinction between core and non-core species is meaningful. In the high-migration regime, we consider the mean residence time of all species τ. In steady state, entry and exit rates of all species into the system must be balanced, $\mu = R^*/\tau$. Using our estimate of R^*(22) gives

$$\tau \approx \frac{-N}{1-\mu} \ln \mu. \tag{33}$$

5. DISCUSSION

In this paper, we proposed a neutral model for cooperative ecosystems. As in the classic neutral theory of biodiversity [34], the strength of our approach lies in its minimalist design. Our analysis was divided into two parts, focusing on steady-state and dynamical properties, respectively.

In the first part, we derived an expression for the steady-state Simpson diversity index, which allowed us to classify the system into different regimes depending on the migration probability. At high migration rates, species abundances are very low and frequency-dependent effects become negligible. In this regime, we recover the predictions of Hubbell's neutral theory of biodiversity, including a Logseries species abundance distribution in steady state. In contrast, at low migration rates, our model exhibits frequency-dependent reproduction rates that result from cooperative interactions. In this regime, our model predicts the emergence of a bimodal abundance distribution (see Fig. 2), which cannot be derived from the classic neutral theory.

Bimodality allows us to define a core of cooperators, defined as the set of species that belong to the high-abundance component. We derive scaling laws for the core, showing that the effective number of species scales roughly as $N^{1/2}$ with population size and $(-\ln \mu)^{-1/2}$ with migration rate. Due to the presence of the core, the system preserves diversity (maintains low λ^*) even at exponentially small migration rates. Hence, cooperative interactions can dramatically increase ecosystem stability.

Several studies have reported bimodal abundance distributions in gut microbiomes, see for example Refs. [30, Fig. 1], [40, Fig. 1], and [41, Fig. 2]. Previous research has attributed bimodality to emergent niche partitioning [42], intrinsic bistability [43], or a combination of multiple processes [41]. However, our theory suggests a neutral mechanism that generates bimodality under very minimal assumptions, offering a novel explanation of this phenomenon.

In the second part of our paper, we studied the stochastic dynamics of species as they enter and leave the system. In the low-migration regime, the distribution of maximum abundances reached again exhibits a bimodal shape, allowing us to classify species into two types. This classification captures the dynamical signature of the cooperator core, as discussed in the steady-state analysis in the first part of this paper. We showed that species that enter the core achieve much higher abundances and longer residence times than species that do not enter the core.

Experimental validation of our predictions may be plausible using engineered microbial strains. For instance, by creating a synthetic consortium in which each species lacks the ability to synthesize one essential molecule, but is able to acquire it through cross-feeding with any other species. Such syntrophic communities have been successfully engineered, ranging from two-species [44, 45] to many [46].

Further extensions of the theory should consider the heterogeneity of interactions, complex network topologies, and spatial effects. Another interesting direction would be to build a connection to the theory of autocatalytic chemical reaction networks, as studied in research on the origin of life. In particular, our model can be interpreted as a stochastic "hypercycle" [47, 48], i.e., a network of cross-catalytic polymers (or other chemical species) that replicate cooperatively. In this context, our results suggest a possible mechanism for the spontaneous emergence of stable and long-lived autocatalytic cores in chemical systems.

ACKNOWLEDGMENTS

JP and RS thank the hospitality of the Santa Fe Institute, where this work started. JP also thanks the hospitality of the Simon Levin Lab and its members, where some of this work was developed. RS acknowledges support from the AGAUR 2021 SGR 0075 grant and the Santa Fe Institute. This project

was partly supported by Grant No. 62417 from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. AK was partly supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 101068029. Finally, SR thanks NSF grant DMR-1910736 for financial support.

[1] J. L. Bronstein, *Mutualism*. Oxford University Press, 2015.
[2] E. G. Leigh Jr, "The evolution of mutualism," *Journal of evolutionary biology*, vol. 23, no. 12, pp. 2507–2528, 2010.
[3] F. Lutzoni and M. Pagel, "Accelerated evolution as a consequence of transitions to mutualism," *Proceedings of the National Academy of Sciences*, vol. 94, no. 21, pp. 11422–11427, 1997.
[4] P. Schuster, "How does complexity arise in evolution?," *Evolution of Complex Systems*, vol. 22, 2001.
[5] J. Bascompte and P. Jordano, "Plant-animal mutualistic networks: the architecture of biodiversity," *Annu. Rev. Ecol. Evol. Syst.*, vol. 38, no. 1, pp. 567–593, 2007.
[6] S. Suweis, F. Simini, J. R. Banavar, and A. Maritan, "Emergence of structural and dynamical properties of ecological mutualistic networks," *Nature*, vol. 500, no. 7463, pp. 449–452, 2013.
[7] M. J. Müller, B. I. Neugeboren, D. R. Nelson, and A. W. Murray, "Genetic drift opposes mutualism during spatial population expansion," *Proceedings of the National Academy of Sciences*, vol. 111, no. 3, pp. 1037–1042, 2014.
[8] S. Germerodt, K. Bohl, A. Lück, S. Pande, A. Schröter, C. Kaleta, S. Schuster, and C. Kost, "Pervasive selection for cooperative cross-feeding in bacterial communities," *PLoS computational biology*, vol. 12, no. 6, p. e1004986, 2016.
[9] E. J. Culp and A. L. Goodman, "Cross-feeding in the gut microbiome: ecology and mechanisms," *Cell host & microbe*, vol. 31, no. 4, pp. 485–499, 2023.
[10] K. Zengler and L. S. Zaramela, "The social network of microorganisms—how auxotrophies shape complex communities," *Nature Reviews Microbiology*, vol. 16, no. 6, pp. 383–390, 2018.
[11] K. Z. Coyte, J. Schluter, and K. R. Foster, "The ecology of the microbiome: networks, competition, and stability," *Science*, vol. 350, no. 6261, pp. 663–666, 2015.
[12] B. E. Morris, R. Henneberger, H. Huber, and C. Moissl-Eichinger, "Microbial syntrophy: interaction for the common good," *FEMS microbiology reviews*, vol. 37, no. 3, pp. 384–406, 2013.
[13] S. Saavedra, F. Reed-Tsochas, and B. Uzzi, "A simple model of bipartite cooperation for ecological and organizational networks," *Nature*, vol. 457, no. 7228, pp. 463–466, 2009.
[14] R. M. May, *Stability and complexity in model ecosystems*. Princeton university press, 2019.
[15] S. Allesina and S. Tang, "Stability criteria for complex ecosystems," *Nature*, vol. 483, no. 7388, pp. 205–208, 2012.
[16] S. P. Hubbell, *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Princeton University Press, 2011.
[17] R. V. Solé, D. Alonso, and A. McKane, "Self–organized instability in complex ecosystems," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 357, no. 1421, pp. 667–681, 2002.
[18] U. Bastolla, M. A. Fortuna, A. Pascual-García, A. Ferrera, B. Luque, and J. Bascompte, "The architecture of mutualistic networks minimizes competition and increases biodiversity," *Nature*, vol. 458, no. 7241, pp. 1018–1020, 2009.
[19] F. S. Valdovinos, "Mutualistic networks: moving closer to a predictive theory," *Ecology letters*, vol. 22, no. 9, pp. 1517–1534, 2019.
[20] P. R. Guimaraes Jr, "The structure of ecological networks across levels of organization," *Annual Review of Ecology, Evolution, and Systematics*, vol. 51, no. 1, pp. 433–460, 2020.
[21] I. Volkov, J. R. Banavar, S. P. Hubbell, and A. Maritan, "Neutral theory and relative species abundance in ecology," *Nature*, vol. 424, no. 6952, pp. 1035–1037, 2003.
[22] J. Harte, *Maximum entropy and ecology: a theory of abundance, distribution, and energetics*. OUP Oxford, 2011.
[23] S. Azaele, S. Suweis, J. Grilli, I. Volkov, J. R. Banavar, and A. Maritan, "Statistical mechanics of ecological systems: Neutral theory and beyond," *Reviews of Modern Physics*, vol. 88, no. 3, p. 035003, 2016.
[24] J. Grilli, "Macroecological laws describe variation and diversity in microbial communities," *Nature communications*, vol. 11, no. 1, p. 4743, 2020.
[25] J. P. O'Dwyer and J. L. Green, "Field theory for biogeography: a spatially explicit model for predicting patterns of biodiversity," *Ecology letters*, vol. 13, no. 1, pp. 87–95, 2010.
[26] R. Muneepeerakul, E. Bertuzzo, H. J. Lynch, W. F. Fagan, A. Rinaldo, and I. Rodriguez-Iturbe, "Neutral metacommunity models predict fish diversity patterns in mississippi–missouri basin," *Nature*, vol. 453, no. 7192, pp. 220–222, 2008.
[27] J. Rosindell, S. P. Hubbell, and R. S. Etienne, "The unified neutral theory of biodiversity and biogeography at age ten," *Trends in ecology & evolution*, vol. 26, no. 7, pp. 340–348, 2011.
[28] A. T. Neu, E. E. Allen, and K. Roy, "Defining and quantifying the core microbiome: challenges and prospects," *Proceedings of the National Academy of Sciences*, vol. 118, no. 51, p. e2104429118, 2021.
[29] G. Wu, T. Xu, N. Zhao, Y. Y. Lam, X. Ding, D. Wei, J. Fan, Y. Shi, X. Li, M. Li, *et al.*, "A core microbiome signature as an indicator of health," *Cell*, vol. 187, no. 23, pp. 6550–6565, 2024.
[30] L. Lahti, J. Salojärvi, A. Salonen, M. Scheffer, and W. M. De Vos, "Tipping elements in the human intestinal ecosystem," *Nature communications*, vol. 5, no. 1, p. 4344, 2014.
[31] E. Simpson, "Measurement of diversity," *Nature*, vol. 163, 1949.
[32] The Simpson index $\lambda(\mathbf{n})$ is sometimes alternatively defined as $\Sigma_{i=1}^{S} \frac{n_i(n_i-1)}{N(N-1)}$, i.e., as the probability of drawing two individuals of the same species when sampling without replacement. The definition (4) corresponds to the same probability when sampling with replacement. The two definitions differ by a small term of order $1/(N-1)$, which is irrelevant for our analysis.
[33] A. E. Magurran, *Measuring biological diversity*. John Wiley & Sons, 2003.
[34] S. P. Hubbell, "A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs," *Coral reefs*, vol. 16, pp. S9–S21, 1997.
[35] M. Malek-Mansour and G. Nicolis, "A master equation description of local fluctuations," *Journal of Statistical Physics*, vol. 13, pp. 197–217, Sept. 1975.
[36] A. McKane, D. Alonso, and R. V. Solé, "Mean-field stochastic theory for species-rich assembled communities," *Physical Review E*, vol. 62, no. 6, p. 8466, 2000.

[37] M. R. Zaghloul, "Efficient multiple-precision computation of the scaled complementary error function and the dawson integral," *Numerical Algorithms*, vol. 95, no. 3, pp. 1291–1308, 2024.
[38] S. G. Johnson, "Faddeeva W function implementation," *URL http://ab-initio. mit. edu/Faddeeva*, 2012.
[39] B. Ottino-Löffler and M. Kardar, "Population extinction on a random fitness seascape," *Phys. Rev. E*, vol. 102, p. 052106, Nov 2020.
[40] M. Loftus, S. A.-D. Hassouneh, and S. Yooseph, "Bacterial associations in the healthy human gut microbiome across populations," *Scientific reports*, vol. 11, no. 1, p. 2828, 2021.
[41] P. I. Costea, F. Hildebrand, M. Arumugam, F. Bäckhed, M. J. Blaser, F. D. Bushman, W. M. De Vos, S. D. Ehrlich, C. M. Fraser, M. Hattori, *et al.*, "Enterotypes in the landscape of gut microbial community composition," *Nature microbiology*, vol. 3, no. 1, pp. 8–16, 2018.
[42] R. Vergnon, E. H. Van Nes, and M. Scheffer, "Emergent neutrality leads to multimodal species abundance distributions," *Nature communications*, vol. 3, no. 1, p. 663, 2012.
[43] D. Gonze, L. Lahti, J. Raes, and K. Faust, "Multi-stability and the origin of microbial community types," *The ISME journal*, vol. 11, no. 10, pp. 2159–2166, 2017.
[44] W. Shou, S. Ram, and J. M. Vilar, "Synthetic cooperation in engineered yeast populations," *Proceedings of the National Academy of Sciences*, vol. 104, no. 6, pp. 1877–1882, 2007.
[45] D. R. Amor, R. Montañez, S. Duran-Nebreda, and R. Solé, "Spatial dynamics of synthetic microbial mutualists and their parasites," *PLOS Computational Biology*, vol. 13, no. 8, p. e1005689, 2017.
[46] M. T. Mee, J. J. Collins, G. M. Church, and H. H. Wang, "Syntrophic exchange in synthetic microbial communities," *Proceedings of the National Academy of Sciences*, vol. 111, no. 20, pp. E2149–E2156, 2014.
[47] M. Eigen and P. Schuster, *The hypercycle: a principle of natural self-organization*. Springer Science & Business Media, 2012.
[48] E. Szathmáry and L. Demeter, "Group selection of early replicators and the origin of life," *Journal of theoretical biology*, vol. 128, no. 4, pp. 463–486, 1987.
[49] P. S. Laplace, *Traité de mecanique céleste, Tome 4*. 1805.
[50] K. B. Oldham, J. Myland, and J. Spanier, *An atlas of functions: with equator, the atlas function calculator*. Springer, 2009.

Supplemental Material: Neutral theory of cooperators

Jordi Piñero, Artemy Kolchinsky, Sidney Redner and Ricard Solé

CONTENTS

1. NEUTRAL MODEL OF COOPERATORS

A. Setup and notation

Our system consists of a well-mixed population of N individuals, each belonging to a certain species. Throughout, we use $i, j, \ldots$ as species indices and use $n_i \in \{0, 1 \ldots, N\}$ to denote the number of individuals belonging to species i. We use bold symbols to denote population vectors, such as $\mathbf{n} = (n_1, n_2, \ldots)$ for the total number of individuals for each species. Recalling the notation used in the main text, we denote by $R(\mathbf{n})$ the number of nonzero elements of the vector $\mathbf{n}$, that is, the number of species present in the system. Moreover, if any species becomes extinct ($n_i \to 0$), the species are re-indexed so that the first $R(\mathbf{n})$ entries of $\mathbf{n}$ are always the strictly positive ones.

Due to our hypothesis of neutrality, all species evolve according to the same rules. Without loss of generality, we take $i = 1$ to be the *representative species*, which, by neutrality, has the same statistical properties as any other species. It will be useful to define

$$\mathbf{n}^\circ := (n_2, n_3, \ldots) \tag{S1.1}$$

as the vector of individual counts for the *non-representative* species. In order to alleviate notation, we interchangeably use $n_1 \leftrightarrow n$ whenever appropriate.

A central concept of our analysis is the Simpson index, defined as

$$\lambda(\mathbf{n}) = \frac{\sum_i n_i^2}{\left(\sum_i n_i\right)^2} = \sum_i \left(\frac{n_i}{N}\right)^2. \tag{S1.2}$$

where we used the fact that $\sum_i n_i = N$. As shown below, we also use the Simpson index for the non-representative species,

$$\lambda(\mathbf{n}^\circ) = \frac{\sum_{i>1} n_i^2}{\left(\sum_{i>1} n_i\right)^2} = \sum_{i>1} \left(\frac{n_i}{N - n_1}\right)^2 \tag{S1.3}$$

Finally, we denote the steady-state expectation of the Simpson index of all species as λ^*, and the Simpson index of the nonrepresentative species as $\lambda_\circ^*$.

B. Birth-death process describing the representative species

In this subsection, we derive a birth-death process (one-dimensional random walk) that describes the stochastic dynamics of n, the abundance of the representative species.

Box I.1: Birth probability. In a single timestep, the probability that the representative species undergoes the transition $n \to n+1$ is

$$(1-\mu)\left(\frac{n}{N}\right)\left(\frac{N-n}{N}\right)\left(\frac{N-n}{N}\right). \tag{S1.4}$$

Here, $1-\mu$ is the probability that the timestep involves a replication (rather than migration), n/N is the probability that an individual from the representative species is chosen to replicate, $(N-n)/N$ is the probability that this individual is paired with an individual from a different species, and the final $(N-n)/N$ is the probability that the offspring replaces an individual of a different species. Since every migration event introduces a new species into the system, there is no contribution to the birth probability of the representative species due to migration.

Following from rules 1) and 2) in the main text, we derive the birth and death transition probabilities (see Boxes 1.1 and 1.2 for a full derivation) as

$$b_n = (1-\mu)\left(\frac{n}{N}\right)\left(1-\frac{n}{N}\right)^2 \tag{S1.5}$$

$$d_{n|\mathbf{n}^\circ} = (1-\mu)\left(\frac{n}{N}\right)\left(1-\frac{n}{N}\right)\left[1-\left(1-\frac{n}{N}\right)\lambda(\mathbf{n}^\circ)\right]+\mu\left(\frac{n}{N}\right) \tag{S1.6}$$

The birth-death process described by (S1.5)-(S1.6) is not closed. In particular, the death probability $d_{n|\mathbf{n}^\circ}$ depends on the Simpson index of the non-representative species, $\lambda(\mathbf{n}^\circ)$. Assuming this Simpson index is tightly concentrated on its steady-state expectation value, $\lambda(\mathbf{n}^\circ) \approx \lambda_\circ^* := \langle \lambda_\circ(\mathbf{n}) \rangle$, we may approximate steady-state fluctuations by a closed birth-death process:

$$b_n = (1-\mu)\left(\frac{n}{N}\right)\left(1-\frac{n}{N}\right)^2 \tag{S1.7}$$

$$d_n = (1-\mu)\left(\frac{n}{N}\right)\left(1-\frac{n}{N}\right)\left[1-\left(1-\frac{n}{N}\right)\lambda_\circ^*\right]+\mu\left(\frac{n}{N}\right). \tag{S1.8}$$

Box I.2: Death transition probability. We first consider the probability that, in a given timestep, the representative species undergoes the transition $n \to n-1$ due to a cooperative interaction. Let B and C denote two species different from each other and from A (the representative species). Then, we may have

$$(B+C)+A \to (B+C)+B \quad \text{or} \quad (B+A)+A \to (B+A)+B,$$

where parentheses denote cooperative interaction. Adding up all possible contributions gives

$$\begin{aligned}
&\sum_{i\neq 1}\left(\frac{n_i}{N}\right)\left(\frac{N-n-n_i}{N}\right)\left(\frac{n}{N}\right)+\left(\frac{N-n}{N}\right)\left(\frac{n}{N}\right)\left(\frac{n}{N}\right)\\
&=\left(\frac{N-n}{N}\right)\left(\frac{n}{N}\right)\sum_{i\neq 1}\left(\frac{n_i}{N}\right)-\left(\frac{n}{N}\right)\sum_{i\neq 1}\left(\frac{n_i}{N}\right)\left(\frac{n_i}{N}\right)+\left(\frac{n}{N}\right)\left(\frac{N-n}{N}\right)\left(\frac{n}{N}\right)\\
&=\left(\frac{n}{N}\right)\left[\left(\frac{N-n}{N}\right)\left(\frac{N-n}{N}\right)-\sum_{i\neq 1}\left(\frac{n_i}{N}\right)\left(\frac{n_i}{N}\right)+\left(\frac{N-n}{N}\right)\left(\frac{n}{N}\right)\right]\\
&=\left(\frac{n}{N}\right)\left(\frac{N-n}{N}\right)\left[1-\sum_{i\neq 1}\left(\frac{n_i}{N-n}\right)\left(\frac{n_i}{N-1}\right)\right]\\
&=\left(\frac{n}{N}\right)\left(\frac{N-n}{N}\right)\left[1-\left(\frac{N-n}{N}\right)\lambda(\mathbf{n}^{\circ})\right],
\end{aligned} \tag{S1.9}$$

where in the last line we used definition (S1.3). In the death transition probability (S1.6), this sum is weighed by $1-\mu$, the probability that a replication event happens in a given timestep. In addition, the death probability includes the term $\mu(n/N)$, representing the probability μ that an individual of a new species enters the system times the probability n/N that this new migrant replaces an individual of the representative species.

2. STEADY-STATE SPECIES ABUNDANCE DISTRIBUTION

In the steady state of a one-dimensional random walk, the detailed balance condition must hold, meaning that birth and death fluxes are balanced:

$$P_n b_n = P_{n+1} d_{n+1}, \quad \text{for } n \in \{1,\dots,N-1\}. \tag{S2.10}$$

Rearranging, we have that

$$P_n = \frac{b_{n-1}}{d_n}P_{n-1} = \cdots = P_1 \prod_{k=1}^{n-1}\frac{b_k}{d_{k+1}} = P_1 \frac{\prod_{k=1}^{n-1} b_k}{\prod_{k=2}^{n} d_k}. \tag{S2.11}$$

Using the birth probabilities, Eq. (S1.7), we may write the numerator on the right-hand side in (S2.11) as

$$\begin{aligned}\prod_{k=1}^{n-1} b_k &= \prod_{k=1}^{n-1}(1-\mu)\frac{k}{N}\frac{N-k}{N}\frac{N-k}{N} \\ &= \left(\frac{1-\mu}{N^3}\right)^{n-1}(n-1)!\frac{(N-1)!}{(N-n)!}\frac{(N-1)!}{(N-n)!}\end{aligned} \tag{S2.12}$$

For the death probabilities derived in (S1.8), it is useful to approximate

$$\begin{aligned}d_k &\approx \frac{k}{N}\left[\frac{N-k}{N}\left(1-\frac{N-k}{N}\lambda_\circ^*\right)-\mu\frac{N-k}{N}+\mu\right] \\ &= \frac{k}{N}\left[\frac{N-k}{N}\left(1-\frac{N-k}{N}\lambda_\circ^*\right)+\mu\frac{k}{N}\right] \\ &\approx \frac{k}{N}\frac{N-k}{N}\left(1-\frac{N-k}{N}\lambda_\circ^*\right) \\ &= \frac{\lambda_\circ^*}{N^3}k(N-k)\left(N/\lambda_\circ^*-N+k\right).\end{aligned} \tag{S2.13}$$

In the first line of (S2.13), we assumed that $\mu\lambda_\circ^* \ll 1$, which means that at large migration, the system becomes very diverse, and, at small migration, the system retains sufficient diversity. In the second line of (S2.13), we assumed that the system is concentrated on small relative abundances when μ is large, thus $\mu(k/N)^2 \ll 1$. Then, the product in the denominator of the right-hand side in (S2.11) is approximated as

$$\prod_{k=2}^{n} d_k \approx \left(\frac{\lambda_\circ^*}{N^3}\right)^{n-1} n!\frac{(N-2)!}{(N-n-1)!}\frac{\Gamma\left(N/\lambda_\circ^*-N+n+1\right)}{\Gamma\left(N/\lambda_\circ^*-N+2\right)}, \tag{S2.14}$$

where Γ indicates Euler's gamma function.

Combining results, and ignoring the factors independent of n (which contribute only to the overall normalization constant), we now compute the ratio in (S2.11) as

$$\begin{aligned}P_n &\propto \frac{(1-\mu)^n}{n(N-n)}\frac{\lambda_\circ^{*-n}}{(N-n)!\,\Gamma\left(N/\lambda_\circ^*-N+n+1\right)} \\ &= \frac{\left(\frac{1-\mu}{1-\lambda_\circ^*}\right)^n}{n(N-n)}\frac{(1-\lambda_\circ^*)^n\lambda_\circ^{*-n}}{(N-n)!\,\Gamma\left(N/\lambda_\circ^*-N+n+1\right)}\end{aligned} \tag{S2.15}$$

To relate our results to a known probability distribution, we multiply the previous expression by $(1-\lambda_\circ^*)^{N/\lambda_\circ^*-N}$ and $\Gamma(N/\lambda_\circ^*+1)$, which are both independent of n, such that

$$P_n \propto \frac{\left(\frac{1-\mu}{1-\lambda_\circ^*}\right)^n}{n(N-n)}\frac{\Gamma(N/\lambda_\circ^*+1)}{(N-n)!\,\Gamma\left(N/\lambda_\circ^*-N+n+1\right)}(1-\lambda_\circ^*)^{N/\lambda_\circ^*-N+n}\lambda_\circ^{*N-n}. \tag{S2.16}$$

This way, expression (S2.16) consists of a combination of distributions involving the well-known binomial distribution, though defined using Euler$-\Gamma$ functions instead of factorials:

$$P_n \propto \frac{(1-\mu)^n}{n(N-n)}\frac{\mathrm{Bin}(N-n;N/\lambda_\circ^*,\lambda_\circ^*)}{(1-\lambda_\circ^*)^n}, \tag{S2.17}$$

where we use notation:

$$\text{Bin}(m; M, p) := \binom{M}{m} p^m (1-p)^{M-m} \equiv \frac{\Gamma(M+1)}{\Gamma(m+1)\Gamma(M-m+1)} p^m (1-p)^{M-m}. \tag{S2.18}$$

In general, $\text{Bin}(m; M, p)$ refers to binomial distribution over $m \in \{0, \ldots, M\}$ successes, given M total trials and probability of success p.

In most cases, we are interested in systems with non-vanishing migration and non-negligible diversity ($\mu > 0$ and $1/\lambda_\circ^* \gg 1$). In such cases, the steady-state distribution is concentrated on small relative abundances and we may rewrite (S2.17) by introducing four approximations. First, we estimate the binomial distribution as a discretized Gaussian (ignoring overall constants):

$$\text{Bin}(N-n; N/\lambda_\circ^*, \lambda_\circ^*) \,\widetilde{\propto}\, e^{-n^2/2N}. \tag{S2.19}$$

Second, when $\lambda_\circ^* \ll 1$, we approximate the exponential term $(1-\lambda_\circ^*)^{-n} \approx e^{n\lambda_\circ^*}$. Third, we take $1/(N-n) \approx 1/N$, which is accurate for $n \ll N$. Finally, we simplify by replacing $\lambda_\circ^*$ (the Simpson index of non-representative species) by the expected Simpson index of the entire population:

$$\lambda_\circ^* \approx \lambda^* := \langle \lambda(\mathbf{n}) \rangle \tag{S2.20}$$

where $\langle \cdot \rangle$ indicates expectations under the steady-state distribution of abundances $\mathbf{n}$. Combining the above approximations, we are left with the probability distribution reported in Eq. (10) in the main text.

To justify (S2.20), let us consider the definitions (S1.2)-(S1.3), and observe that

$$\lambda(\mathbf{n}) = \left(\frac{N}{N-n}\right)^2 \lambda(\mathbf{n}^\circ) + \left(\frac{n}{N}\right)^2 \geq \lambda(\mathbf{n}^\circ). \tag{S2.21}$$

For the upper bound, we may rewrite the equation as

$$\begin{aligned} \lambda(\mathbf{n}) &= \lambda(\mathbf{n}^\circ) + \frac{2Nn - n^2}{(N-n)^2} \lambda(\mathbf{n}^\circ) + \left(\frac{n}{N}\right)^2 \\ &\leq \lambda(\mathbf{n}^\circ) + \frac{2n/N}{(1-n/N)^2} + \left(\frac{n}{N}\right)^2 \end{aligned} \tag{S2.22}$$

where in the last line we used that $n^2 \geq 0$ and $\lambda(\mathbf{n}^\circ) \leq 1$. Observe that $\langle n \rangle / N$ is the inverse of the number of non-extinct species, see Eq. (20) in the main text. Thus, assuming that there is a large number of species in the system, the steady-state probability distribution should be concentrated on $n \ll N$. Then, taking expectations of both sides and ignoring terms of order n/N gives $\lambda^* := \langle \lambda(\mathbf{n}) \rangle \approx \lambda_\circ^* := \langle \lambda_\circ(\mathbf{n}) \rangle$.

3. STEADY-STATE SIMPSON INDEX

Here we show that the steady-state Simpson index of the system obeys Eq. (11) and can be approximated by the expression (12) given in the main text. To begin, we rewrite Eq. (S1.2) as

$$\lambda(\mathbf{n}) \approx \sum_{n=1}^{N} r_n \left(\frac{n}{N}\right)^2 \tag{S3.23}$$

where r_n indicates the number of species with abundance n. Observe that $\sum_n r_n n = N$, thus $r_n n/N$ is the *fraction* of individuals that belong to the set of species with abundance n. The steady-state expectation of r_n is proportional to the probability that the representative species has abundance n, that is, $\langle r_n \rangle \propto P_n$. The expected fraction of individuals that belong to the set of species with abundance n is obtained by normalizing as

$$\frac{\langle r_n \rangle n}{N} = \frac{P_n n}{\sum_{n'} P_{n'} n'}. \tag{S3.24}$$

As a sanity check, note that $\sum_n \langle r_n \rangle n/N = \langle \sum_n r_n n \rangle /N = 1$. Finally, let us take expectation on both sides of (S3.23) and substitute to give

$$\lambda^* = \langle \lambda(\mathbf{n}) \rangle \approx \sum_{n=1}^{N} \left(\frac{\langle r_n \rangle n}{N}\right)\left(\frac{n}{N}\right) = \frac{1}{N}\frac{\sum_{n=1}^{N} P_n n^2}{\sum_{n'=1}^{N} P_{n'} n'}, \tag{S3.25}$$

which corresponds to Eq. (11) in the main text. Usefully, the right side is independent of the normalization constant of P_n, which is not always easy to compute in practice.

A. Low migration Simpson index

In the low-migration regime, we can approximate (S3.25) via the Euler-Maclaurin integrals. First, we rewrite our expression for P_n, Eq. (10), as

$$P_n \propto \frac{1}{n} e^{-(n-N\lambda^*)^2/2N + n\ln(1-\mu)} \approx \frac{1}{n} e^{-[n-N(\lambda^*-\mu)]^2/2N}, \tag{S3.26}$$

where we used that, for small migration values, $\ln(1-\mu) \approx -\mu$ and ignored overall constant factors. Then, combining with (S3.25), we obtain

$$\lambda^* \approx \frac{1}{N} \frac{\int_1^N e^{-[n-N(\lambda^*-\mu)]^2/2N} n\, dn}{\int_1^N e^{-[n'-N(\lambda^*-\mu)]^2/2N} dn'}. \tag{S3.27}$$

The two integrals can be solved separately. Assuming $N \gg 1$,

$$\begin{aligned} \frac{1}{N}\int_1^N e^{-[n-N(\lambda^*-\mu)]^2/2N} n\, dn &\approx e^{-N(\lambda^*-\mu)^2/2} + \sqrt{\frac{\pi N}{2}}(\lambda^*-\mu)\,\mathrm{erfc}\left(\sqrt{\frac{N}{2}}(\mu-\lambda^*)\right) \\ \int_1^N e^{-[n'-N(\lambda^*-\mu)]^2/2N} dn' &\approx \sqrt{\frac{\pi N}{2}}\mathrm{erfc}\left(\sqrt{\frac{N}{2}}(\mu-\lambda^*)\right), \end{aligned} \tag{S3.28}$$

where we used the complementary error function, $\mathrm{erfc}(z) := 2/\sqrt{\pi}\int_z^\infty e^{-t^2}dt$. Combining with (S3.27) yields

$$\lambda^* \approx \left[\sqrt{\frac{\pi N}{2}}\mathsf{erfcx}\left(\sqrt{\frac{N}{2}}(\mu-\lambda^*)\right)\right]^{-1} + (\lambda^*-\mu), \tag{S3.29}$$

where we used the definition of the scaled complementary error function, $\mathsf{erfcx}(z) := e^{z^2}\mathrm{erfc}(z)$. Rearranging (S3.29), we arrive at the result reported in Eq. (12) from the main text.

B. High migration Simpson index derived from (12)

In the main text, the scaling $\lambda^* \approx 1/N\mu$ in the high-migration regime ($\mu \gg \lambda$) is obtained by re-evaluating (11) using the Fisher Logseries distribution. In contrast, here we show that the same result can also be derived from (12). To do so, we may approximate

$$\mathsf{erfcx}(z) \approx z/[\sqrt{\pi}(z^2+1/2)] \quad \text{for} \quad z\to\infty, \tag{S3.30}$$

as follows by considering the first two terms of the continued fraction derived by Laplace [49, Livre X, p. 255]. Plugging into (12) and solving gives

$$\lambda^* \approx \frac{\mu-\sqrt{\mu^2-4/N}}{2} \approx \frac{1}{N\mu} \tag{S3.31}$$

where we used our usual assumption that N is large.

4. STEADY-STATE NUMBER OF SPECIES

Here, we derive the expected number of species in steady state, distinguishing between those in the cooperator core, R^*_{core}, and those outside of it, R^*_{out}. Our derivation is only valid in the low-migration regime introduced and discussed in Sections 3-3 A in the main text. The results derived here appear as Eqs. (24) and (25) in the main text.

As stated in the main text, the expected number of species in the core obeys $R^*_{\mathsf{core}} = N_{\mathsf{core}}/\langle n\rangle_{\mathsf{core}}$. We may use $N_{\mathsf{core}} + N_{\mathsf{out}} = N$ to write

$$N_{\mathsf{core}} = N\left(1-\frac{N_{\mathsf{out}}}{N}\right). \tag{S4.32}$$

We estimate the fraction of individuals outside the core by

$$\frac{N_{\mathsf{out}}}{N} = \frac{\sum_{n=1}^{\hat{n}_{\mathsf{min}}} P_n n}{\sum_{n=1}^{N} P_n n} \approx \frac{\sum_{n=1}^{1/\lambda^*} e^{-(n-N\lambda^*)^2/2N}}{\sum_{n=1}^{N} e^{-(n-N\lambda^*)^2/2N}} \approx \frac{\int_1^{1/\lambda^*} e^{-(n-N\lambda^*)^2/2N}}{\int_1^{N} e^{-(n-N\lambda^*)^2/2N}}, \tag{S4.33}$$

where we used that $N_{\mathsf{out}} = \sum_{n=1}^{\hat{n}_{\mathsf{min}}} R^* P_n n$ and $N = \sum_{n=1}^{N} R^* P_n n$, where R^* is the total number of species. We also use that $\hat{n}_{\mathsf{min}} \approx 1/\lambda^*$, see Eq. (15) in the main text. We note that the denominator on the right side acts as the

normalization constant of the Gaussian probability density function (PDF), which now together with the numerator gives the Gaussian cumulative density function (CDF), often indicated as $\Phi(z)$. Thus, in the large N limit, we write:

$$\frac{N_{\mathsf{out}}}{N} \approx \Phi\left(\frac{1/\lambda^* - N\lambda^*}{\sqrt{N}}\right) \tag{S4.34}$$

As shown in Eq. (13), for very small migration, $\lambda^* \gg N^{-1/2}$, which allows us to approximate

$$\Phi\left(\frac{1/\lambda^* - N\lambda^*}{\sqrt{N}}\right) \approx \Phi\left(-N\lambda^*\right) \approx \Phi(-\infty) = 0.$$

Therefore, to a first approximation, we have that

$$N_{\mathsf{core}} = N\left(1 - \frac{N_{\mathsf{out}}}{N}\right) \approx N. \tag{S4.35}$$

Finally, we derive Eq. (24) by combining with $R^*_{\mathsf{core}} = N_{\mathsf{core}}/\langle n\rangle_{\mathsf{core}}$ and using the approximation $\langle n\rangle_{\mathsf{core}} \approx \hat{n}_{\mathsf{max}}$, with the last term specified in Eq. (15).

Regarding the expected number of species outside the core, we similarly use $R^*_{\mathsf{out}} = N_{\mathsf{out}}/\langle n\rangle_{\mathsf{out}}$. To proceed, we start by estimating the expected abundance of species outside the core,

$$\langle n\rangle_{\mathsf{out}} = \frac{\sum_{n=1}^{\hat{n}_{\mathsf{min}}} P_n n}{\sum_{n=1}^{\hat{n}_{\mathsf{min}}} P_n} \approx \frac{N_{\mathsf{out}}}{N}\frac{\sum_{n=1}^{N} P_n n}{\sum_{n=1}^{1/\lambda^*} P_n}. \tag{S4.36}$$

where, again, we used that $N_{\mathsf{out}} = \sum_{n=1}^{\hat{n}_{\mathsf{min}}} R^* P_n n$ and $N = \sum_{n=1}^{N} R^* P_n n$, where R^* is the total number of species. We also use that $\hat{n}_{\mathsf{min}} \approx 1/\lambda^*$, see Eq. (15) in the main text. Rearranging, and substituting P_n by using our main result Eq. (10) from the main text, gives

$$R^*_{\mathsf{out}} = \frac{N_{\mathsf{out}}}{\langle n\rangle_{\mathsf{out}}} \approx \frac{N\sum_{n=1}^{1/\lambda^*} e^{-(n-N\lambda^*)^2/2N}/n}{\sum_{n=1}^{N} e^{-(n-N\lambda^*)^2/2N}}, \tag{S4.37}$$

We estimate the sum of the denominator in the right-hand side of (S4.37) by the Gaussian normalization factor, $\sqrt{2\pi N}$. For the numerator, we approximate

$$\sum_{n=1}^{1/\lambda^*} \frac{1}{n} e^{-(n-N\lambda^*)^2/2N} \approx e^{-(1-N\lambda^*)^2/2N} \sum_{n=1}^{1/\lambda^*} \frac{1}{n} \approx -\sqrt{2\pi N}\mu \ln \lambda^*, \tag{S4.38}$$

where in the last approximation we evaluated the (slow-changing) exponential term in the range $1 \leq n \leq 1/\lambda^*$ as $e^{-(1-N\lambda^*)^2/2N} \approx e^{N{\lambda^*}^2/2}$, and then plugged in our estimation of λ^* in the low-migration regime (Eq. (12) in the main text). Combining the above gives (25) as reported in the main text.

5. MAXIMUM ABUNDANCES AND THE INFILTRATION PROBABILITY

A. Maximum abundances distribution

We begin by deriving the probability that a species trajectory (sampled at stationarity) has maximum abundance m, which we term Q_m. First, we consider the probability that an abundance trajectory with initial value n reaches (or exceeds) abundance $\ell \geq n$ ($\ell \neq 0$). We term this probability $u_n(\ell)$. Second, we derive an analytic expression for $u_1(\ell)$, the probability of reaching (or exceeding) ℓ when the trajectory starts at abundance one. Third, we connect results for u_1 to Q_m and study the shape of Q_m in the low-migration regime. And fourth, we estimate the *infiltration probability*, i.e., the probability that a trajectory reaches maximum abundance $m \geq N\lambda^*$.

Consider the probability that a trajectory with initial abundance n reaches (or exceeds) some target abundance $\ell \geq n$ ($\ell \neq 0$) before going extinct. We denote this probability as $u_n(\ell)$. We focus in particular on $u_1(\ell)$, which captures the probability that a species that has just migrated into the system reaches (or exceeds) abundance value $\ell > 1$. We will focus our analysis on the regime of low migration.

As described above, our model can be reduced to a birth-death process with transition probabilities b_n and d_n, Eqs. (S1.7)-(S1.8). We are also using the approximation $\lambda^*(\mathbf{n}) \approx \lambda^*$. Then, the probability $u_n(\ell)$ must obey two boundary conditions:

$$u_0(\ell > 0) = 0 \quad \text{and} \quad u_\ell(\ell) = 1. \tag{S5.39}$$

The first condition states that the probability of hitting a target once the species has gone extinct ($n \to 0$) is null, while the second condition states that the probability of hitting the target while on target is one. For the values of $n \in \{1, \ldots, \ell - 1\}$, the probability $u_n(\ell)$ obeys a recurrence:

$$u_n(\ell) = b_n u_{n+1}(\ell) + d_n u_{n-1}(\ell) + (1 - b_n - d_n) u_n(\ell)\,. \tag{S5.40}$$

We rearrange (S5.40) to give:

$$b_n[u_{n+1}(\ell) - u_n(\ell)] = d_n[u_n(\ell) - u_{n-1}(\ell)]. \tag{S5.41}$$

We then define the difference $w_n(\ell) = u_n(\ell) - u_{n-1}(\ell)$, which obeys $w_1(\ell) = u_1(\ell)$ since $u_0(\ell) = 0$. Then, from (S5.41), we write

$$w_n(\ell) = \frac{d_{n-1}}{b_{n-1}} w_{n-1}(\ell) = \prod_{k=1}^{n-1} \frac{d_k}{b_k} w_1(\ell). \tag{S5.42}$$

We note that this product is similar to the inverse product obtained in (S2.11), but differs by one in the indexing of

the death probability factors. Moreover, since $u_n(\ell) = \sum_{m=1}^{n} w_m(\ell)$, the recurrence (S5.40) is solved by

$$u_n(\ell) = \sum_{m=1}^{n} \prod_{k=1}^{n-1} \frac{d_k}{b_k} u_1(\ell). \tag{S5.43}$$

We use the condition $u_\ell(\ell) = 1$ to write

$$u_1(\ell) = \left[\sum_{m=1}^{\ell} \prod_{k=1}^{m-1} \frac{d_k}{b_k} \right]^{-1}, \tag{S5.44}$$

which is an exact enumeration for $u_1(\ell)$.

To find a closed-form approximation of Eq. (S5.44) for our model, we recall results (S2.12), (S2.13) and (S2.14), and carefully rearrange the indices where needed. After some algebra, this results in

$$\prod_{k=1}^{m-1} \frac{d_k}{b_k} \approx \left(\frac{\lambda^*}{1-\mu} \right)^{m-1} \frac{(N-m)!}{(N-1)!} \frac{\Gamma(N/\lambda^* - N + m)}{\Gamma(N/\lambda^* - N + 1)}. \tag{S5.45}$$

Next, we use the Stirling approximation $\Gamma(z+1) \approx \sqrt{2\pi} z^{z+1/2} e^{-z}$ (valid for $z \gg 1$), and apply it to each of the terms in (S5.45). Expanding terms up to second order in m/N, assuming $N \gg 1$ and $\lambda^* \ll 1$, we reach

$$\prod_{k=1}^{m-1} \frac{d_k}{b_k} \approx \exp\left\{ \frac{m^2}{2N} - \left(\lambda^* + \ln(1-\mu) + \frac{1}{2N} \right) m + \lambda^* + \ln(1-\mu) \right\}. \tag{S5.46}$$

We now ignore the term $1/2N$ and use the fact that, in the low-migration regime, $\lambda^* \gg \mu$. This allows us to approximate (S5.46) as

$$\prod_{k=1}^{m-1} \left(\frac{d_k}{b_k} \right) \approx e^{m^2/2N - \lambda^*(m-1)}. \tag{S5.47}$$

We now plug (S5.47) back into (S5.44) and for convenience introduce the notation $x_\ell := (\ell - N\lambda^*)/\sqrt{2N}$. This gives

$$\frac{1}{u_1(\ell)} \approx \int_1^{\ell} \prod_{k=1}^{m-1} \left(\frac{d_k}{b_k} \right) dm \approx e^{\ell^2/2N - (\ell-1)\lambda^*} \left[\sqrt{2N}\mathscr{D}(x_\ell) + \frac{1}{2} \right] - \left[\sqrt{2N}\mathscr{D}(x_1) - \frac{1}{2} \right], \tag{S5.48}$$

where we used the Euler-Maclaurin approximation of the sum as an integral plus boundary terms, and ignored terms of order $O(1/N)$ in the exponents. We also introduced the Dawson function:

$$\mathscr{D}(z) := e^{-z^2} \int_0^z e^{t^2} dt\,. \tag{S5.49}$$

After a bit of rearranging, we arrive at Eq. (28), the expression in $u_1(\ell)$ presented in the main text.

B. Bimodality of Q_m

It can be seen from Fig. 6 (inset) in the main text that the distribution Q_m has a bimodal shape, with a local minimum that separates two modes. This allows us to assign species trajectories into two classes: those whose maximum

abundance value crosses over the threshold located at the local minimum of Q_m and those that do not. We denote the first set as the *dynamical core*.

To identify the location of the local minimum of Q_m, recall from the main text that $Q_m = u_1(m) - u_1(m+1)$, thus we can approximate $Q_m \approx -du_1/dm$. Then, the local minimum is given by the condition

$$0 = \frac{d}{dm}Q_m \approx -\frac{d^2}{dm^2}u_1(m) \propto -2\left(\frac{d}{dm}v_m\right)^2 + v(m)\frac{d^2}{dm^2}v_m \tag{S5.50}$$

where, for notational convenience, we introduced $v_m = 1/u_1(m)$ to denote the right side of Eq. (S5.48). Considering that equation, it can be shown that the derivatives of v_m become very small at $m = N\lambda^*$, where the minimum of the exponent's argument is reached. Linearizing the right side of Eq. (S5.50) around $m \approx N\lambda^*$ and then solving gives

$$m^* \approx N\lambda^* - \frac{1}{2} - \frac{2N - 3/4}{e^{N\lambda^{*2}/2-\lambda^*}\left(\sqrt{2N}\mathscr{D}(x_1) - 1/2\right) + 1}. \tag{S5.51}$$

In Eq. (S5.56), we show that $\sqrt{2N}\mathscr{D}(x_1) \approx 1/\lambda^*$. Since $\lambda^* \gg 1/\sqrt{N}$ in the low-migration regime, the denominator in the last term in Eq. (S5.51) is exponentially large, allowing us to ignore this whole term. We also have $N\lambda^* \gg 1/2$, thus dropping all sub-dominant terms gives the approximation

$$m^* \approx N\lambda^* . \tag{S5.52}$$

To summarize, we have shown that the maximum abundance distribution Q_m has a local minimum at values of $m \approx N\lambda^*$. Interestingly, the value of this local minimum in Q_m corresponds to the center of the Gaussian contribution to the abundance distribution P_n (Eq. (12) in the main text) in the low-migration regime. However, this point is not exactly the high-abundance mode of the steady-state distribution, which is located at $\hat{n}_{\mathsf{max}} = N\lambda^* - 1/\lambda^*$.

C. Infiltration probability

Here we derive the infiltration probability to the dynamical core, β. We will use the asymptotic expansion of the Dawson function [50, 42:6:4],

$$\mathscr{D}(z) = 1/(2z) + O(z^{-3}) \qquad \text{for} \qquad |z| \to \infty . \tag{S5.53}$$

Observe that β corresponds to the cumulative probability to cross over the local minimum of Q_m located at $N\lambda^*$:

$$\beta = \sum_{m=N\lambda^*}^{N} Q_m = \sum_{m=N\lambda^*}^{N} [u_1(m) - u_1(m+1)] = u_1(N\lambda^*) - u_1(N) \approx u_1(N\lambda^*) . \tag{S5.54}$$

Here we first used (26) from the main text, and then $u_1(N) \approx 0$. The latter approximation is justified because $u_1(N) \sim e^{-N/2}$ up to polynomial factors, as follows from Eq. (S5.48) and the fact that the Dawson function

contributes only polynomial terms, see (S5.53). To obtain a closed-form expression for β, we substitute back into (S5.48) and keep leading-order terms to arrive at

$$\beta \approx \frac{1}{-\sqrt{2N}\mathscr{D}(x_1)}\,. \tag{S5.55}$$

Since $x_1 = (1 - N\lambda^*)/\sqrt{2N} \approx -\lambda^*\sqrt{N/2} \ll -1$ in the low-migration regime, Eq. (S5.53) implies

$$\sqrt{2N}\mathscr{D}(x_1) \approx \sqrt{2N}\mathscr{D}\Big(-\lambda^*\sqrt{N/2}\Big) \approx -\sqrt{2N}/(\sqrt{2N}\lambda^*) = -1/\lambda^*\,, \tag{S5.56}$$

allowing us to approximate $\beta \approx \lambda^*$, and thus recovering expression (29) from the main text.

6. RESIDENCE TIMES

In this section, we calculate the mean residence time of species whose trajectories reach the dynamical core, as defined in SM5. We also argue that the cooperator core obtained in our study of the steady-state abundance distribution and the dynamical core are effectively equivalent.

At stationarity, let us denote by T_n the mean first passage time (MFPT) from a species with initial abundance n to reach extinction (at abundance 0). We observe that a core species will spend a long time at high abundance values; hence, we assume that the mean residence of a species whose trajectory belongs to the dynamical core can be approximated by the MFPT from $N\lambda^*$ to extinction,

$$\tau_{\mathsf{core}} \approx T_{N\lambda^*}. \tag{S6.57}$$

The MFPT T_n obeys the following recurrence:

$$T_n = b_n T_{n+1} + d_n T_{n-1} + (1 - b_n - d_n)T_n + 1. \tag{S6.58}$$

By defining the differences $v_n := T_n - T_{n-1}$, this can be rearranged into a first-order recurrence,

$$v_n = \frac{d_{n-1}}{b_{n-1}} v_{n-1} - \frac{1}{b_{n-1}}. \tag{S6.59}$$

Recurrence (S6.59) is generally solved by

$$v_n = U_{n-1}v_1 - Q_{n-1}, \quad \text{with} \quad U_{m-1} := \prod_{k=1}^{m-1} \frac{d_k}{b_k} \quad \text{and} \quad Q_{m-1} = \sum_{k=1}^{m-1} \frac{1}{b_k} \prod_{j=k+1}^{m-1} \frac{d_j}{b_j}. \tag{S6.60}$$

Next, from Eq. (S5.47) above,

$$U_{m-1} \approx e^{m^2/2N - \lambda^*(m-1)}. \tag{S6.61}$$

For the other coefficient, we have

$$Q_{m-1} = \sum_{k=1}^{m-1} \frac{1}{b_k} \prod_{j=k+1}^{m-1} \frac{d_j}{b_j} = U_{m-1} \sum_{k=1}^{m-1} \frac{1}{b_k} \prod_{j=1}^{k} \frac{b_j}{d_j}. \tag{S6.62}$$

The expression for Q_{m-1} is related to the steady-state distribution P_k because

$$P_k d_k = P_{k-1} b_{k-1} \Rightarrow P_k = P_1 \prod_{j=1}^{k-1} \frac{b_j}{d_{j+1}} = d_1 P_1 \frac{1}{b_k} \prod_{j=1}^{k} \frac{b_j}{d_j}, \tag{S6.63}$$

therefore we may rewrite Q_{m-1} as

$$Q_{m-1} = \frac{U_{m-1}}{P_1 d_1} \sum_{k=1}^{m-1} P_k = \frac{U_{m-1}}{P_1 d_1} C(m-1). \tag{S6.64}$$

Here, we defined $C(\ell) = \sum_{k=1}^{\ell} P_k$, as the cumulative distribution of P_k up to ℓ. Returning now to (S6.60),

$$v_n = U_{n-1} \left(v_1 - \frac{1}{P_1 d_1} C(n-1) \right). \tag{S6.65}$$

At this point, we introduce the boundary conditions of the recurrence (S6.58). First, we have $T_0 = 0$, which implies $v_1 = T_1 - T_0 = T_1$. In addition, we have

$$T_N = d_N T_{N-1} + (1 - d_N) T_N + 1 \Leftrightarrow d_N (T_N - T_{N-1}) = 1 \Leftrightarrow v_N = \frac{1}{d_N}. \tag{S6.66}$$

The expected number of species in steady-state, R^*, satisfies the balance equation between outflow due to extinction and inflow due to migration:

$$\mu = R^* P_1 d_1, \tag{S6.67}$$

Recall $d_N = \mu$ from Eq. (6) in the main text. Put together, these results imply that

$$\frac{1}{\mu} = v_N = U_{N-1} \left(T_1 - \frac{R^*}{\mu} C(N-1) \right), \tag{S6.68}$$

which allows us to obtain

$$T_1 = \frac{1}{\mu} \left[\frac{1}{U_{N-1}} + R^* C(N-1) \right] \approx \frac{R^*}{\mu}. \tag{S6.69}$$

Here, we used that, in the low-migration regime, $U_{N-1} \approx e^{N/2 - \lambda^* N} \gg 1$ in the large N limit, while $C(N-1) \approx 1$ since all the mass is concentrated below $n = N$. Therefore,

$$v_n \approx \frac{R^* U_{n-1}}{\mu} (1 - C(n-1)), \tag{S6.70}$$

Using the definition of v_n, the condition $v_1 = T_1$, and setting $n = N\lambda^*$, which is the starting abundance value for our MFPT, then we have

$$T_{N\lambda^*} = \sum_{m=1}^{N\lambda^*} v_m \approx \frac{R^*}{\mu} \sum_{m=1}^{N\lambda^*} U_{m-1} \left[1 - C(m-1) \right]. \tag{S6.71}$$

Next, we observe that in the low-migration regime, P_k is well-approximated as a Gaussian distribution with mean $N\lambda^*$ and standard deviation $\sqrt{N}$ (see Eq. (10) in the main text). Hence, C is approximated as the cumulative density function of this Gaussian distribution, so $1 - C(m-1) \approx 1$ for small m and $1 - C(m-1) \approx 1/2$ for $m \approx N\lambda^*$. On the other hand, for our range of $m \in [1, N\lambda^*] \cap \mathbb{N}$, U_{m-1} has a local maximum at $m = 1$ and, at small values of m, decays exponentially with rate $1/\lambda^*$, see Eq. (S6.61). Therefore, in the low-migration regime where $N\lambda^{*2} \gg 1$, one can verify that as m increases, U_{m-1} decays essentially to 0 while $1 - C(m-1)$ remains very close to 1. This allows us to approximate:

$$\sum_{m=1}^{N\lambda^*} U_{m-1}\left[1 - C(m-1)\right] \approx \sum_{m=1}^{N\lambda^*} U_{m-1}. \tag{S6.72}$$

The right-hand side corresponds to the expression for $1/u_1(N\lambda^*)$ that we approximated in (S5.48). Using our definition of β (S5.54) and our result (S5.55) in terms of the Dawson function, we combine with (S6.71) to give

$$T_{N\lambda^*} \approx \frac{R^*}{\mu}\sqrt{2N}\mathscr{F}_N\left(\frac{N\lambda^* - 1}{\sqrt{2N}}\right) \approx \frac{R^*}{\mu\beta}. \tag{S6.73}$$

Recall that in the low-migration regime, $R^* \approx R^*_{\mathsf{core}}$, which, combined with Eq. (24) in the main text, approximates the MFPT (S6.73) as

$$T_{N\lambda^*} \approx \frac{1}{\mu\lambda^*[1 - (N\lambda^*)^{-1}]}. \tag{S6.74}$$

Result (S6.74) is equal to our estimate of τ_{core} obtained in Eq. (30) in the main text. This shows that the same expression for τ_{core} can be derived using two different arguments: a dynamical argument based on the MFPT calculation, and a steady-state argument based on the principle of detailed balance given in Sec. 4 in the main text.