,

Demographic noise can reverse the direction of deterministic selection

George W. A. Constable,[1] Tim Rogers,[2] Alan J. McKane,[3] and Corina E. Tarnita[1]

[1] *Department of Ecology and Evolutionary Biology, Princeton University, NJ 08544, U.S.A.*
[2] *Department of Mathematical Sciences, University of Bath, BA2 7AY, U.K.*
[3] *School of Physics and Astronomy, The University of Manchester, M13 9PL, U.K.*

Deterministic evolutionary theory robustly predicts that populations displaying altruistic behaviors will be driven to extinction by mutant cheats that absorb common benefits but do not themselves contribute. Here we show that when demographic stochasticity is accounted for, selection can in fact act in the reverse direction to that predicted deterministically, instead favoring cooperative behaviors that appreciably increase the carrying capacity of the population. Populations that exist in larger numbers experience a selective advantage by being more stochastically robust to invasions than smaller populations, and this advantage can persist even in the presence of reproductive costs. We investigate this general effect in the specific context of public goods production and find conditions for stochastic selection reversal leading to the success of public good producers. This insight, developed here analytically, is missed by both the deterministic analysis as well as standard game theoretic models that enforce a fixed population size. The effect is found to be amplified by space; in this scenario we find that selection reversal occurs within biologically reasonable parameter regimes for microbial populations. Beyond the public good problem, we formulate a general mathematical framework for models that may exhibit stochastic selection reversal. In this context, we describe a stochastic analogue to $r - K$ theory, by which small populations can evolve to higher densities in the absence of disturbance.

Over the past century, mathematical biology has provided a framework with which to begin to understand the complexities of evolution. Historically, development has focused on deterministic models [1]. However, when it comes to questions of invasion and migration in ecological systems, it is widely acknowledged that stochastic effects may be paramount, since the incoming number of individuals is typically small. The importance of demographic (intrinsic) noise has long been argued in population genetics; it is the driver of genetic drift and can undermine the effect of selection in small populations [2, 3]. This concept has also found favor in game theoretic models of evolution which seek to understand how apparently altruistic traits can invade and establish in populations [4]. However, the last decade has seen an increase in the awareness of some of the more exotic and counter-intuitive aspects of demographic noise: it has the capacity to induce cycling of species [5], pattern formation [6, 7], speciation [8] and spontaneous organization in systems that do not display such behavior deterministically.

Here we explore the impact of demographic noise on the direction of selection in interactions between multiple phenotypes or species. Historically, a key obstacle to progress in this area has been the analytical intractability of multidimensional stochastic models. This is particularly apparent when trying to investigate problems related to invasion, where systems are typically far from equilibrium. A promising avenue of analysis has recently become apparent however through stochastic fast-variable elimination [9, 10]. If a system consists of processes that act over very different timescales, it is often possible to eliminate fast-modes, assumed to equilibrate quickly in the multidimensional model, and obtain a reduced dimensional description that is amenable to analysis [11]. This approach has been employed multiple times over the last decade to study a stochastic formulation of the classical Lotka-Volterra competition model for two competing phenotypes/species. In [9, 10, 12–14], such models were analyzed under the assumption that the dynamics regulating the total population size (birth, death and competition) occurred on a much faster timescale than the change in population composition. In particular [9, 10, 12, 13] have shown that it is possible for systems that appear neutral in a deterministic setting to become non-neutral once stochasticity is included. If the two phenotypes have equal deterministic fitness, but one is subject to a larger amount of demographic noise than the other, then the effect of this noise alone can induce a selective drift in favor of the phenotype experiencing less noise. This stems from the fact that it is easier to invade a noisy population than a stable one; furthermore, the direction of this induced selection can vary with the system's state [15]. The idea has been further generalized mathematically in [16].

Here we will show more generally that not only can stochasticity break deterministic neutrality, but that it has the capacity to reverse the direction of selection predicted deterministically. Thus while in a deterministic setting a certain phenotype will always reach fixation (and is resistant to invasions), in a stochastic setting its counterpart can in fact be more likely to invade and fixate (and less susceptible to invasions). These results generalize recent work on modified Moran and Wright-Fisher type models [17, 18] to a large class of models consisting of two phenotypes interacting with their environment. We begin with the analysis of a prototypical public good model, which is used to illustrate our analysis. We find

that stochastic selection reversal can alleviate the public good production dilemma. We further show how space can amplify this phenomenon, allowing the reversal of selection to emerge over a greater parameter range. Finally, we extend the ideas to a more general model framework, and explore the types of system in which we expect this behavior to be relevant. In particular we discuss the similarities with $r-K$ selection theory [19].

I. PUBLIC GOOD MODEL

It is generally accepted that random events play a strong role in the evolution of cooperative behavior, which is deterministically selected against [4]. The standard formulation of evolutionary game theory involves setting the problem in terms of a modified Moran model [20, 21]. The Moran model is a population genetic model first developed as an abstract illustration of the effect of genetic drift in a haploid population of two phenotypes; an individual is picked to reproduce with a probability proportional to their fitness, whilst simultaneously a second individual is chosen randomly to die [22]. Coupling birth and death events keeps the population size fixed, which increases the tractability of the system.

The specification of fixed population size is however restrictive and can be problematic. Most prominently, a phenotype with increased fitness can be no more abundant in isolation than its ailing counterpart. Additional difficulties are encountered if one attempts to use simple game-theoretic models to quantitatively understand more complex experimental data. While, for example, assuming some arbitrary non-linearity in the model's game payoff matrix may enable experimental findings to be elegantly recapitulated, it is more difficult to justify the origin of these assumptions on a mechanistic level [23]. In light of such issues, it has been suggested that a more ecologically grounded take on the dynamics of cooperation might be preferable [24, 25], one in which the population size is not fixed and that is sufficiently detailed that mechanistic (rather than phenomenological) parameters can be inferred experimentally. In the following, we take such an approach. We begin by considering a prototypical model of public good production and consumption.

In our model, we consider a phenotype X having the ability to produce a public good Q that catalyzes its growth. We wish to capture the stochastic dynamics of the system. To this end we assume that the system is described by a set of probability transition rates, which describe the probability per unit time of each reaction occurring:

$$X \underset{\kappa/R^2}{\overset{b_x}{\rightleftarrows}} X+X\,, \quad X+Q \xrightarrow{r/R^2} X+X+Q\,,$$
$$X \xrightarrow{p_x} X+Q\,, \quad Q \xrightarrow{\delta} \varnothing\,. \tag{1}$$

In the absence of the public good, the producer phenotype X reproduces at a baseline birthrate b_x. The phenotypes encounter each other and the public good at a rate R^{-2}; the quantity R^2 can be interpreted as a measure of the area (or volume) to which the system is confined. Death of the phenotype occurs solely due to crowding effects at rate κ, multiplied by the encounter rate. Phenotypes encounter and utilize the public good at a rate r/R^2. We study the case where this reaction is catalytic (i.e. the public good is conserved) and leads to a phenotype reproduction. Examples of catalytic (reusable) public goods are the enzyme invertase produced by the yeast *Saccharomyces cerevisiae* [26] or the siderophore pyoverdine produced by the bacterium *Pseudomonas aeruginosa* [27]. The total rate at which the phenotype reproduces is thus increased in the presence of the public good. The public good itself is produced by the producer phenotype at a rate p_x and decays at a rate δ. Note that as well as controlling the spatial scale of the well-mixed system, the magnitude of R will also control the typical number of individuals in the system, since larger R (more space) allows the population to grow to greater numbers. We next introduce a mutant phenotype Y that does not produce the public good; (i.e. $p_y = 0$) consequently, it has a different baseline birth rate b_y which we expect to be at least as high as that of the producer, due to the non-producers' reduced metabolic expenditure. Its interactions with the public good are otherwise similar to those of X (see Eq. (1)).

The state of the system is specified by the discrete variables n_x, n_y and n_q, the number of each phenotype and public good respectively. For the system described, we wish to know the probability of being in any given state at any given time. To answer this, we set up an infinite set of partial difference equations (one for each unique state (n_x, n_y, n_q)) that measures the flow of probability between neighboring states (controlled by the transitions Eq. (1)). These equations govern the time-evolution of a probability density function $P(n_x, n_y, n_q, t)$ (see Eq. (A2)). Such a model is sometimes termed a microscopic description [28], since it takes account of the dynamics of discrete interactions between the system variables.

Although the probabilistic model is straightforward to formalize, it is difficult to solve in its entirety. We apply an approximation that makes the model more tractable, while maintaining the system's probabilistic nature. Such approximations, which assume that the system under consideration has a large but finite number of individuals, are well practiced and understood [28] and are analogous to the diffusion approximation [22] of population genetics. Assuming that R is large, but finite, (which implies a large number of individuals in the system), we transform the system into the approximately continuous variables $(x, y, q) = (n_x, n_y, n_q)/R^2$ and expand the partial difference equations in $1/R^2$. This allows us to to express the infinite set of partial difference equations as a single partial differential equation in four continuous variables, (x, y, q, t). However, since the

PDE results from a Taylor expansion, it has infinite order. Truncating the expression after the first term (at order R^{-2}), one obtains a deterministic approximation of the dynamics (valid for $R \to \infty$, or equivalently for infinite population sizes). Since we aim to make the system tractable but still retain some stochastic element to the dynamics, we truncate the expansion after the second term (at order R^{-4}, see Eq. (A4)). The resulting model can be conveniently expressed as a set of Itō stochastic differential equations (SDEs):

$$\begin{aligned} \dot{x} &= x\left[b_x + rq - \kappa(x+y)\right] + R^{-1}\eta_x(t)\,, \\ \dot{y} &= y\left[b_y + rq - \kappa(x+y)\right] + R^{-1}\eta_y(t)\,, \\ \dot{q} &= p_x x - \delta q + R^{-1}\eta_q(t)\,. \end{aligned} \tag{2}$$

The $\eta_i(t)$ represent Gaussian white noise terms whose correlations depend on the state of the system (the noise is multiplicative). Importantly, because Eq. (2) has been developed as a rigorous approximation of the underlying stochastic model, Eq. (1), the precise functional form of the noise can be determined explicitly, rather than posited on an ad-hoc basis (see Appendix A). Setting $R \to \infty$, the population size increases with the interaction scale and one recovers the deterministic limit. Since Eq. (2) is a course-grained approximation of the underlying microscopic model but retains an inherent stochasticity, it is often referred to as the mesoscopic limit [29].

First we analyze the dynamics of Eq. (2) in the deterministic, $R \to \infty$ limit. There exist three fixed points, or equilibria. The first, at the origin, is always unstable. The remaining fixed points occur when the system only contains a single phenotype: the producer fixed point, $(x, y, q) = (K_x, 0, p_x K_x/\delta)$ and the non-producer fixed point, $(x, y, q) = (0, K_y, 0)$. Thus K_x and K_y are measures of the phenotypes' frequency (carrying capacity) in isolation, with precise forms

$$K_x = \frac{b_x \delta}{\kappa\delta - p_x r}\,, \qquad K_y = \frac{b_y}{\kappa}\,. \tag{3}$$

If $b_y > b_x$ then the non-producer fixed point is always stable while the producer fixed point is always unstable. However, the non-producer fixed point is only globally attracting if $\kappa\delta > r p_x$. If this condition is not met then there exist initial conditions for which the producers produce and process the public good faster than they die and faster than the public good degrades, resulting in unbounded exponential growth of the system. This biologically unrealistic behavior comes from the fact that we have assumed for simplicity that the public good uptake does not saturate. Since this behavior is unrealistic, we will work in the regime $\kappa\delta > r p_x$ for the remainder of the paper. Finally, we are interested in systems where the size of the producer population in isolation is larger than that of the non-producer, $K_x > K_y$; this is true if the condition $b_x > b_y(1 - r p_x/\delta\kappa)$ holds. Thus, deterministically, a non-producing mutant will always take over a producer population and, due to the absence of the public good, it will yield a smaller population at equilibrium.

This deterministic analysis predicts, unsurprisingly, that a population composed entirely of non-producers is the only stable state. We next explore the behavior of the system in Eq. (1) when demographic stochasticity is considered.

A. Mesoscopic selection reversal

Due to noise, a stochastic system will not be positioned precisely on deterministic fixed points, but rather it will fluctuate around them. In the above system, these fluctuations will occur along the y-axis for the non-producer fixed point while in the absence of non-producers they will occur in the (x, q) plane for the producer fixed point. We can define $N_x = R^2 K_x$ and $N_y = R^2 K_y$ to be the mean number of the phenotypes X and Y in isolation in the respective stationary states. We assume that the non-producing phenotype has a greater per-capita birth rate than the producer phenotype, i.e. $b_y > b_x$, and we introduce a single non-producing mutant into a producer population. While the deterministic theory predicts that the non-producer should sweep through the population until it reaches fixation, in the stochastic setting fixation of the non-producer is by no means guaranteed: there is a high probability that the single mutant might be lost due to demographic noise. However, since the non-producer is deterministically selected for, we might expect the probability of a non-producer mutant invading and fixating in a resident producer population to be greater than the probability of a producer mutant invading and fixating in a resident non-producer population. We will explore this question below.

In order to make analytic predictions about the stochastic model, we need to reduce the complexity of the system. This can be done if we employ methods based on the elimination of fast variables [30] to obtain an effective one-dimensional description of the system dynamics. To this end, we begin by assuming that the public good production and decay, p_x and δ, and the phenotypes' reproduction and death, b_x, b_y, and κ, occur on a much faster timescale than the rate of change of population composition, which is governed by the difference in birth rates, $b_x - b_y$. Essentially this assumes that the cost of public good production is marginal. In the case of *S. cerevisiae*, this assumption is supported by empirical work (see Table S.2). In order to mathematically investigate this timescale-separation we define

$$b_x = b(1-\varepsilon)\,, \quad b_y = b\,, \tag{4}$$

where the parameter ε represents the metabolic cost that X pays for producing the public good. The parameter ε now controls the rate of change of population composition, and if $1 >> \varepsilon$, we have our desired timescale separation in the deterministic system. Because the parameters K_x, K_y, N_x and N_y depend on ε, we will find it convenient to define their values when $\varepsilon = 0$ as $K_x^{(0)}$,

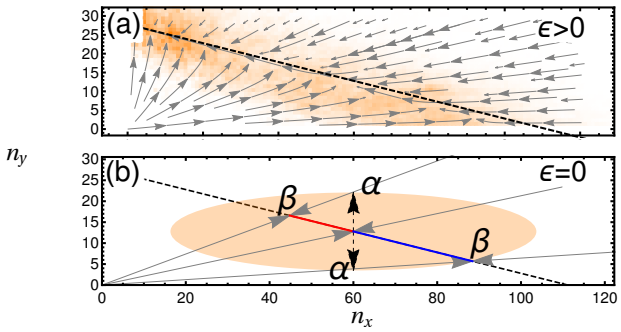


FIG. 1. System dynamics in the phenotype plane. Deterministic trajectories shown as gray arrows. Panel (a): Trajectories rapidly collapse to a SM (black dashed line), before slowly moving to the non-producing Y fixed-point. Stochastic trajectories (histogram overlaid in orange) remain in the region of the SM but may fluctuate away from it. Panel (b): Illustration of the origin of noise-induced selection. The orange ellipse depicts the standard deviation of Gaussian fluctuations originating at its center. Fluctuations (black dashed arrows) to points α are equally likely, however when projected back to the CM (black dashed line) to points β, a bias for producing X phenotype is observed. Parameters used are $p_x = 9.5 \times 10^{-4}$, $\varepsilon = 0.08$ in panel (a), $\varepsilon = 0$ in panel (b) and the remaining parameters are given in Table S.2.

$K_y^{(0)}$, $N_x^{(0)}$ and $N_y^{(0)}$ respectively. In order to maintain our assumption that the composition of the phenotype population changes slowly in the stochastic system, we additionally require that the noise is small. However this assumption has already been implicitly made in the derivation of Eq. (2), where it was assumed that R is large, and thus R^{-1}, the prefactor for the noise terms, is small. In order to formalize this, we will find it convenient to assume $R^{-2} \approx \mathcal{O}(\varepsilon)$.

Under the above assumptions, the system features a separation of timescales. Next, we take advantage of this to reduce the complexity of the system. Deterministically, the existence of a set of fast timescales suggests the existence of a lower-dimensional subspace, the slow manifold (SM), shown in Fig. 1(a), to which the system quickly relaxes, and along which it slowly moves, until it reaches the system's stable fixed point. This behavior can be exploited if we assume that the system reaches the SM instantaneously. We can then describe the dynamics of the entire system in this lower dimensional space, and thus reduce the number of variables in our description of the deterministic system. However, we are interested in the stochastic dynamics.

The stochastic trajectories initially collapse to the region around the SM, about which they are confined, but along which they can move freely until one of the phenotypes fixates (see Fig. 1(a)). Fluctuations that take the system off the SM are quickly quashed back to another point on the SM; however the average position on the SM to which a fluctuation returns is not necessarily the same as that from which the fluctuation originated. A crucial element of the dynamics in this stochastic setting is that the form of the noise, combined with that of the trajectories back to the SM, can induce a bias in the dynamics along the SM (see Fig. 1(b)). This is the origin of the stochastic selection reversal that we will explore. In order to capture this behavior while simultaneously removing the fast timescales in the stochastic system, we map all fluctuations off the SM along deterministic trajectories back to the SM [30]. This essentially assumes that any noisy event that takes the system off the SM is instantaneously projected back to another point on the SM.

For clarity, we briefly describe the dynamics when $\varepsilon = 0$. In this case the birth rates of phenotypes X and Y are identical. Instead of the two non-zero fixed points, K_x and K_y, found above, the deterministic system now has a line of fixed points, referred to as a center manifold (CM) [31]. The CM is identical to the SM in the limit $\varepsilon \to 0$. It is given by

$$y = \frac{K_y^{(0)}}{K_x^{(0)}} \left(K_x^{(0)} - x \right), \quad q = \frac{p_x}{\delta} x, \tag{5}$$

and shown graphically in Fig. 1(b). The separation of timescales in the system is now at its most pronounced, since there are strictly no deterministic dynamics along the CM following the fast transient to the CM. However the stochastic system still features dynamics along the CM. Applying the procedure outlined in [30] we arrive at a description of the stochastic dynamics in a single variable, the frequency of producers along the CM;

$$\dot{x} = \frac{b}{R^2} x \left(1 - \frac{x}{K_x^{(0)}} \right) \mathcal{F}(x) + \frac{1}{R} \zeta(t), \tag{6}$$

where

$$\mathcal{F}(x) = 2 \left(\frac{K_x^{(0)} - K_y^{(0)}}{(K_x^{(0)} K_y^{(0)})^2} \right) \left[K_x^{(0)} K_y^{(0)} + \left(K_x^{(0)} - K_y^{(0)} \right) x \right].$$

Here $\zeta(t)$ is a Gaussian white noise term with a correlation structure given in Eq. (B10). Together with Eq. (5), Eq. (6) approximates the dynamics of the entire system. Note that while Eq. (6) predicts a noise-induced directional drift along the CM (controlled by $\mathcal{F}(x)$), a deterministic analysis predicts no dynamics, since the CM is by definition a line of fixed points. This directional drift along the CM results from the projection bias illustrated in Fig. 1(b). If $p_x > 0$, then $K_x^{(0)} > K_y^{(0)}$, and so $\mathcal{F}(x) > 0$; thus the public good production by phenotype X induces a selective pressure that selects for X along the center manifold.

The origin of the term $\mathcal{F}(x)$ in Eq. (6) can be understood more fully by exploring its implications for the invasion probabilities of X and Y, denoted ϕ_x and ϕ_y. These can be straightforwardly calculated since the system is one-dimensional (see Appendix C). We find

$$\phi_x = \frac{1}{N_y}, \quad \text{and} \quad \phi_y = \frac{1}{N_x}, \tag{7}$$

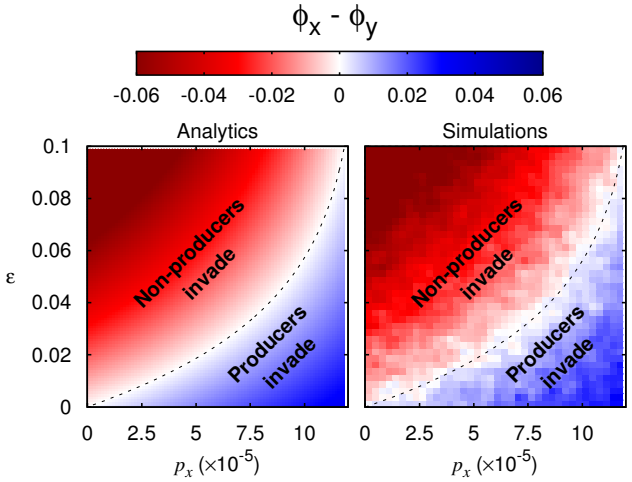


FIG. 2. Stochasticity can render non-producers more susceptible to invasion by producers than vice versa. Plots of the difference in invasion probabilities between producers X and non-producers Y as a function of the cost to birth for production, ε, and good production rate p_x. The remaining parameters are taken from Table S.2. Left: analytic results for a single small patch (see Eq. (9)). The critical cost ε for selection reversal, Eq. (9), is shown here as the black dashed line. Right; results from Gillespie simulations [46] of the stochastic process Eq. (1), averaged over 2000 runs.

where $\phi_x > \phi_y$ so long as $p_x > 0$ (see Eq. (3)). The term $\mathcal{F}(x)$ can thus be interpreted as resulting from the stochastic advantage the producers have at the population level from reaching higher carrying capacities in isolation, which makes them more stochastically robust to invasion attempts. This result is independent of the spatial scale R (and therefore population size) so long as R is finite.

If $\varepsilon \neq 0$, the system does not collapse to the CM, but rather to the SM. At leading order in ε, the equation for the SM is given by Eq. (5). Upon removing the fast dynamics, the effective dynamics of x can now be shown to take the form (see Eq. (B14))

$$\dot{x} = bx\left(1 - \frac{x}{K_x^{(0)}}\right)\left(\frac{1}{R^2}\mathcal{F}(x) - \varepsilon\right) + \frac{1}{R}\zeta(t)\,, \qquad (8)$$

where $\zeta(t)$ and $\mathcal{F}(x)$ are the same as in Eq. (6). The SDE now consists of two components. The deterministic contribution, governed by ε, exerts a selective pressure against phenotype X, due to its reduced birth rate. The stochastic term, $\mathcal{F}(x)$ exerts a pressure in favor of phenotype X, resulting, as in the case $\varepsilon = 0$ discussed above, from the producers stochastic robustness to invasions.

Thus, when $\varepsilon > 0$, a trade-off emerges in the stochastic system between the stochastic advantage to public good production (due to increased population sizes) and the deterministic cost producers pay (in terms of birth rates). If the birth costs are not too high, producers will be selected for, which constitutes a reversal in the direction of selection from the deterministic prediction. Specifically, we can calculate the condition on the metabolic cost that ensures that the producers are fitter than the non-producers (i.e. $\phi_x > \phi_y$):

$$\varepsilon < \frac{\kappa}{bR^2}\log\left[\frac{\delta\kappa}{\delta\kappa - p_x r}\right]. \qquad (9)$$

Whereas for no metabolic cost producers consistently have a stochastic advantage regardless of typical population size (see Eq. (7)), for non-zero production costs, the population must be sufficiently small that stochastic effects, governed by R^{-2}, are dominant. Fig. 2 shows that the theory predicts well the trade-off in the underlying stochastic model (1).

We have shown that stochastic selection reversal is more prevalent when R is not large. Meanwhile our analytic results results have been obtained under the assumption that R is large, which allowed us to utilize the diffusion approximation leading to Eq. (2) and aided the timescale elimination procedure that yielded Eq. (8). We therefore expect that although stochastic selection reversal will become more prominent as R is reduced, the quality of our analytic predictions may suffer. Despite this caveat, it is the small R regime that is interesting to us. Small values of R are associated with small population sizes. While it is conceivable that populations of macro-organisms may consist of a small number of individuals, this limit is not so pertinent to the study of micro-organisms. In the next section however, we will show that by incorporating space, the constraint of small population size can be relaxed.

II. SPATIAL AMPLIFICATION

In this section we consider a metapopulation on a grid: each subpopulation (patch) has a small size so that demographic noise continues to be relevant locally, but the number of subpopulations is large so that the overall population in the system is large. This method of incorporating demographic stochasticity into spatial systems has proved to be successful in the modeling of microbial populations [7]. We consider a grid of C patches. The dynamics within each patch are given by the transitions in Eq. (1), and coupled to the surrounding patches by the movement of the phenotypes and public good. A patch will produce migrants at a rate proportional to its density. Producers X and non-producers Y disperse with a probability rate m to a surrounding region, while the public good diffuses into neighboring regions at a rate D. Once again the diffusion approximation can be applied to obtain a set of SDEs approximating the system dynamics;

$$\begin{aligned}
\frac{\mathrm{d}x_{ij}}{\mathrm{d}\tau} &= x_{ij}\left(b_x + rq_{ij} - \kappa(x_{ij}+y_{ij})\right) + m\left(L\boldsymbol{x}\right)_{ij} + \frac{\eta_{xij}(t)}{R}\,,\\
\frac{\mathrm{d}y_{ij}}{\mathrm{d}\tau} &= y_{ij}\left(b_y + rq_{ij} - \kappa(x_{ij}+y_{ij})\right) + m\left(L\boldsymbol{y}\right)_{ij} + \frac{\eta_{yij}(t)}{R}\,,\\
\frac{\mathrm{d}q_{ij}}{\mathrm{d}\tau} &= p_x x_{ij} - \delta q + D\left(L\boldsymbol{q}\right)_{ij} + \frac{\eta_{qij}(t)}{R}\,, \qquad (10)
\end{aligned}$$

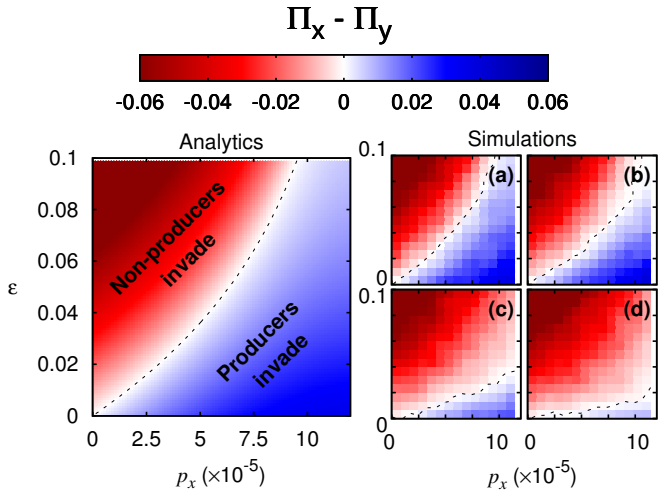


FIG. 3. Left panel: analytic results show that space amplifies stochastic selection reversal in the low dispersal, zero diffusion limit. The critical maximum cost ε for selection reversal (see Eq. (15)) is plotted as black dashed line. Right panel: simulation results are shown for varied m and D, averaged over 2000 runs. Panel (a) gives results which in the low dispersal, zero diffusion limit ($m = 3.7 \times 10^{-8}$, $D = 0$), which match our theoretical predictions. Panel (b) gives the result with a set of biologically plausible parameters ($m = 3.7 \times 10^{-7}$ and $D = 2.2 \times 10^{-5}$ derived in Table S.2). Panel (c) gives results in a system with high dispersal ($m = 3.7 \times 10^{-5}$, $D = 2.2 \times 10^{-5}$), while panel (d) shows the case of a system with high diffusion ($m = 3.7 \times 10^{-7}$ and $D = 2.2 \times 10^{-3}$). The number of patches is given by $C = 16$ and the remaining parameters are listed in Table S.2.

where ij is the patch on row i and column j. The operator L is the discrete Laplacian operator $(L\boldsymbol{x})_{ij} = -4x_{ij} + x_{(i-1)j} + x_{(i+1)j} + x_{i(j-1)} + x_{i(j+1)}$. If $b_y > b_x$, the deterministic dynamics predict that the producers will always go extinct.

First we will discuss some important limit case behavior for this system. In the limit of large dispersal rate m and diffusion rate D, the stochastic system behaves like a well-mixed population with a spatial scale cR^2 (i.e. the spatial structure is lost). In this case, as the size of the spatial system is increased, the effective population size also increases, and as a consequence selection reversal for producing phenotypes becomes less likely (see Eq. (9)).

We next consider the low-dispersal, zero diffusion limit. For sufficiently low dispersal, any incoming mutant will first either fixate or go to extinction locally before any further dispersal event occurs. Since each dispersal/invasion/extinction event resolves quickly, at the population level, the system behaves like a Moran process on a graph [4], with each node representing a patch. The 'fitness' of a patch is the probability that it produces a migrant, and that that migrant successfully invades a homogeneous patch of the opposite type, following the approach used in [17]. Denoting the 'fitness' of producing and non-producing patches by W_x and W_y respectively, we have

$$W_x = mN_x\phi_x\,, \qquad W_y = mN_y\phi_y\,, \tag{11}$$

where N_i $(i = x, y)$ is the mean carrying capacity of phenotype i in a homogeneous patch, and ϕ_i are the invasion probabilities of a type i mutant in a type $j \neq i$ patch. The fixation probabilities of a homogeneous patch in a population of the opposite phenotype can now be calculated using standard results [4]. Let ρ_i $(i = x, y)$ denote the probability that type i takes over the metapopulation when starting from one patch of type i in a population otherwise comprised entirely of patches of the opposite phenotype. Then

$$\rho_i = \frac{1 - r_i^{-1}}{1 - r_i^{-C}}\,, \quad \text{for} \quad i = x, y \quad \text{and} \quad r_x = \frac{W_x}{W_y}\,, \quad r_y = \frac{W_y}{W_x}\,. \tag{12}$$

If we start from a single invading mutant, the probability that it takes over the entire population (i.e. invasion probability) is the product between the probability that it takes over its home patch ϕ_i, and the probability that the newly invaded home patch fixates into the metapopulation, ρ_i:

$$\Pi_x = \phi_x\rho_x\,, \quad \Pi_y = \phi_y\rho_y\,. \tag{13}$$

In the infinite patch limit ($C \to \infty$), ρ_x and ρ_y depend on r_x, the patch fitness ratio defined in Eq. (12). If $r_x > 1$, $\rho_x \to 1 - r_x^{-1}$ and $\rho_y \to 0$, whereas if $r_x < 1$ the converse is true. This means that, in the infinite patch, low dispersal, zero diffusion limit, the condition for the stochastic reversal of selection is weakened from $\phi_x > \phi_y$ to

$$N_x\phi_x > N_y\phi_y\,. \tag{14}$$

Spatial structure therefore has the ability to enhance the stochastic reversal observed in the small well-mixed system. An approximate analytic form for the above condition can be obtained in terms of the original parameters;

$$\varepsilon < 2\frac{\kappa}{bR^2}\log\left[\frac{\delta\kappa}{\delta\kappa - p_x r}\right]\,. \tag{15}$$

Once again, our analytical results are well supported by simulations (see Fig. 3). The critical production rate for the invasion probability of producers to exceed that of non-producers has been decreased, as predicted by Eqs. (9) and (15). Producers can therefore withstand higher production costs in spatially structured environments.

It is important to note that while Eq. (14) depends on the mean number of producers and non-producers on a homogeneous patch (N_x and N_y), it is independent of the number of individuals in the entire metapopulation in the large C limit. The interaction between these two spatial scales leads to results that can appear counter-intuitive. Demographic noise, as we have discussed, leads to producing patches being 'more fit' at the patch level

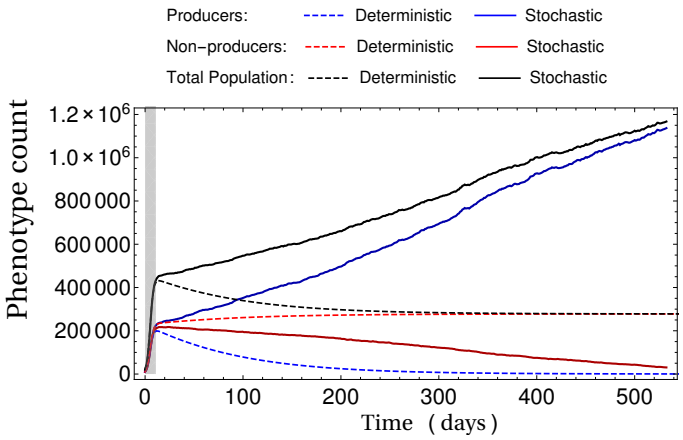


FIG. 4. Demographic stochasticity at the local 'patch' scale profoundly alters the system dynamics at the population level. Results are obtained from stochastic and deterministic ($R \to \infty$) simulations of Eq. (10) with a grid of $C = 100 \times 100$ patches, $p_x = 1 \times 10^{-4}$, $\varepsilon = 0.02$, $m = 3.7 \times 10^{-5}$ and the remaining parameters taken from Table S.2. Initial conditions are a single producer and non-producer on each patch. The initial (fast) transient collapse to the SM occurs occurs in the shaded gray region. Following this, the deterministic system slowly moves along the slow manifold until the non-producers fixate, whereas in the stochastic system, the producers experience a selective pressure in their favor. For dynamics at the patch level, see Supplementary Information movie S1.

(see Eq. (11)). However, when a large number of patches is considered, the demographic noise at the metapopulation level is reduced. This leads to the system following trajectories that appear deterministic at the level of the metapopulation, even though the path they follow is entirely the result of demographic stochasticity at the within-patch level (see Fig. 4). The movie S1 (see Appendix H) displays the individual dynamics of the patches that comprise the trajectory illustrated in Fig. 4.

Away from the small dispersal, zero diffusion limit, the dramatic selection reversal predicted by the analytical results is clearly weakened (see Fig. 3). Though selection reversal is still found across a range of m and D values, if either dispersal or diffusion are too high, the selection reversal breaks down. It is therefore important to understand what order of magnitude estimates for the values of m and D may be biologically reasonable.

A. Insights from *S. cerevisiae*

In the following section, we will attempt to contextualize our model with reference to a *S. cerevisiae* yeast system, which has been previously identified as a biological example of a population that features public good producers and non-producers. The model we have presented is general and therefore it could not capture the full biological detail of this particular system. For instance, it has been noted that some degree of privatization of the public good occurs in even the well-mixed experimental system [23], a behavior we do not consider in our model. However, setting our model in this context can provide some insight into the scenarios in which we might expect stochastic selection reversal to be a biologically relevant phenomenon.

An *S. cerevisiae* yeast cell metabolizes simple sugars, such as glucose, in order to function. However, when simple sugars are scarce, the yeast can produce invertase, an enzyme that breaks down complex sugars, such as sucrose, to release glucose [32]. Invertase is produced at a metabolic cost and, since digestion of sucrose occurs extracellularly, most of the benefits of its production are shared by the population. Specifically in the case of S. cerevisiae, *SUC*2, the wild-type strain, produces invertase, while the lab cultured mutant *suc*2 does not [33]. In terms of our model parameters, the baseline birth rates, b_x and b_y, represent respectively *SUC*2 and *suc*2 reproduction in the absence of invertase. This could be understood as arising from yeast directly metabolizing sucrose (a less energetically beneficial metabolic route [32]) or as the result of some extrinsically imposed low glucose concentration in the system. The rate r would then represent the additional birth rate in the presence of invertase. The form of our specified reactions (see Eq. (1)), assumes that the presence of invertase leads directly to a yeast reproduction event. In reality invertase must break down the sucrose into glucose, and then slowly absorb the glucose. We are therefore essentially assuming that the sucrose is abundant, its breakdown by invertase instantaneous, and the glucose absorption rapid and occurring in discrete packets, with each packet absorbed leading to a reproduction event.

In the well-mixed system, our analytic predictions indicate that stochastic selection reversal can occur only if the population is very small. Since this is an unrealistic assumption in the case of yeast cultures, we would predict that non-producers should come to dominate a well-mixed population. In a spatially structured population however, this constraint is relaxed since it only requires small interaction regions. For *S. cerevisiae*, we can obtain order of magnitude estimates for the majority of parameters in our model, including public good diffusion (see Appendix G). Using these estimates together with our analytic results for the spatial public goods system, we find that stochastic selection reversal could feasibly be an important phenomenon for promoting the evolution of microbial public goods production in spatial settings (see Fig. 3, Panel b). Given this, we now consider the results of a spatial experiment on *S. cerevisiae*, and ask how its results might be interpreted in light of the insights developed with our simple model.

In [33], *SUC*2 and *suc*2 were experimentally competed on an agar plate. It was found that non-producing *suc*2 could not invade from rare (1% of initial yeast population), and in fact decreased in frequency, becoming undetectable at long times (around 800 generations). This suggests that in a spatial setting, invertase producing

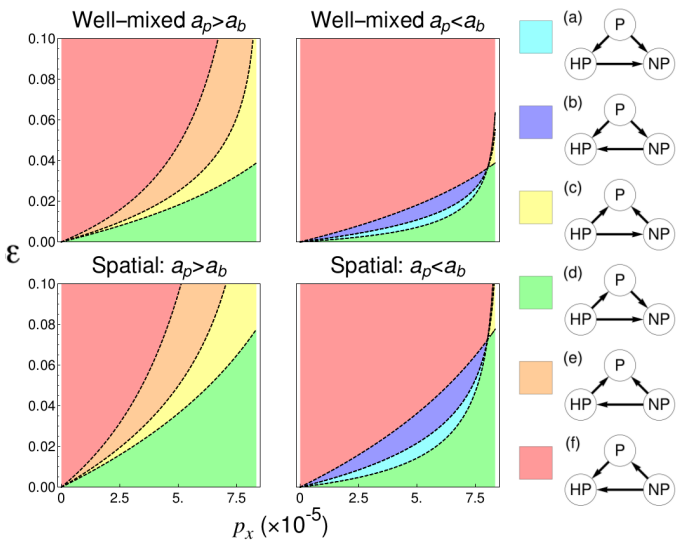


FIG. 5. Plots of the pairwise invariability scenarios possible for non-producing, producing and hyper-producing phenotypes. Arrows point away from the dominant phenotype in a pair, which is defined as that with a larger invasion probability (Fig. S3). Non-transitive dynamics are not possible. It is possible however for an optimal intermediate good production rate to emerge (cyan and blue regions), if $a_p < a_b$. In this scenario the hyper-producer receives diminishing good production as a function of cost to birth rate compared to the producer. Left panels, $a_b = 1.3$ and $a_p = 1.5$. Right panels, $a_b = 3$ and $a_p = 1.5$. Remaining parameters given in Table S.2.

*SUC*2 yeast are robust to invasions, which is in qualitative agreement with our theoretical predictions. The experiments yielded an additional result, the appearance of a hyper-producing mutant. This hyper-producing phenotype produced invertase at approximately 1.5 times the rate of standard producers and existed at higher densities. The hyper-producer appeared to evolve naturally and establish robust colonies during the competition experiments between non-producers and producers. However, when separate competition experiments were conducted between the hyper-producers and the producers, the hyper-producers failed to demonstrate any appreciable fitness advantage over the producers. This potentially suggests an optimal invertase production rate, whereby the hyper-producers managed to establish and grow during the *SUC*2-*suc*2 competition experiments by exploiting non-producing regions due to a relative fitness advantage, but could not invade regions of space occupied by producers. Interestingly, our model also predicts that an intermediate optimal production rate may exist, depending on how the cost of production scales with the production rate. Suppose a hyper-producer, U, produces at a rate $p_u = a_p p_x$, paying a metabolic cost $a_b\varepsilon$ to its birth rate, such that $b_u = b(1 - a_b\varepsilon)$. The pairwise invasion probabilities of each phenotype can then be calculated (see Supplementary Information, Section S.4). We define the fitter phenotype in a pair as that with the larger invasion probability. The potential fitness rankings are investigated in Fig. 5 as a function of p_x and ε, (which we recall also alter p_u and b_u). We draw particular attention to the right panels, in which $a_b > a_p$. In this scenario, the hyper-producers pay a disproportionate cost for their increased production rate compared to the producers. This can be interpreted as diminishing returns for production. In this case, there exist regions where the producer is the optimal phenotype (regions (a) and (b), in blue and cyan respectively). Specifically, scenario (a) displays a similar behavior to that observed in [33], in which producers win out over both non-producers and hyper-producers, but hyper-producers are more likely to invade non-producing populations.

III. GENERALITY OF RESULTS

We have shown that demographic stochasticity can reverse the direction of selection in a public good model. In this section we will show that the mechanism responsible for this phenomenon is by no means particular to this model. We consider a general scenario, with a phenotype X_1, which is at the focus of our study, and a number of discrete ecosystem constituents, E_i. In the public good model for instance, we would label the public good itself as an ecosystem constituent, however more generally this could be a food source, a predator or anything else that interacts with the phenotypes. The state of the ecosystem influences the birth and death of the phenotype and in turn the presence of the phenotype influences the state of the ecosystem, altering the abundances of the constituents. We assume that the system lies at a unique, stable stationary state, precluding the possibility of periodic behavior. Suppose that a new phenotype, X_2, arises. We assume that the second phenotype is only slightly better at exploiting the ecosystem than X_1, though its influence on the ecosystem may be very different. For instance, in the public goods model, non-producers have a small birth-rate advantage over producers, but do not produce the public good. Which phenotype is more likely to invade and fixate in a resident population of the opposite type?

The stochastic model for this system can be constructed in a similar manner to the public good model; the dynamics are described by a set of probability transition rates (analogous to Eq. (1)). We restrict the transitions by specifying that although the two phenotypes compete, there is no reaction that instantaneously changes both of their numbers in the population. This final condition simply means that they should not, for instance, be able to mutate from one type to another during their lifetime, or to prey on each other. A parameter R is introduced, to once again govern the typical scale of the system. The model is analyzed in the mesoscopic limit, by introducing $(x_1, x_2, \boldsymbol{e}) = (n_{x1}, n_{x2}, \boldsymbol{n}_e)/R^2$ and applying the diffusion approximation. For large but finite R,

the mesoscopic description takes the form

$$\begin{aligned} \dot{x_1} &= x_1 F^{(0)}(\boldsymbol{x},\boldsymbol{e}) - \varepsilon x_1 F^{(\varepsilon)}(\boldsymbol{x},\boldsymbol{e}) + R^{-1}\eta_1(t)\,, \\ \dot{x_2} &= x_2 F^{(0)}(\boldsymbol{x},\boldsymbol{e}) + R^{-1}\eta_2(t)\,, \\ \dot{e_i} &= F_i(\boldsymbol{x},\boldsymbol{e}) + R^{-1}\beta_i(t)\,, \quad \forall\, i = 3,\dots J\,, \end{aligned} \tag{16}$$

where ε is small and governs selective pressure against X_1. The assumption that there is no reaction that instantaneously changes the number of both phenotypes ensures that the correlation structure of the noise terms takes the form

$$\begin{aligned} &\langle \eta_1(t)\eta_1(t')\rangle = \delta(t-t')x_1 H^{(0)}(\boldsymbol{x},\boldsymbol{e})\,, \\ &\langle \eta_2(t)\eta_2(t')\rangle = \delta(t-t')x_2 H^{(0)}(\boldsymbol{x},\boldsymbol{e})\,, \quad \langle \eta_1(t)\eta_2(t')\rangle = 0\,, \end{aligned}$$

with ε taken to be of order R^{-2}. This assumption, made here to isolate the effect of varying carrying capacity from any other intraspecies dynamics, means that while the magnitude of fluctuations in the number of both phenotypes is dependent on the state of the system, $(\boldsymbol{x},\boldsymbol{e})$, the fluctuations themselves are not correlated with each other. Restrictions on the microscopic model that yield the above SDE description are addressed more thoroughly in Appendix E. The form of Eq. (16) makes the nature of the system we describe more clear; it consists of two competing phenotypes, which reproduce according to replicator dynamics [1] with equal fitness at leading order in ε.

In the special case $\varepsilon = 0$, both phenotypes are equally fit, regardless of their influence on the ecosystem variables e_i. The degeneracy of the dynamics in x_1 and x_2 ensures the existence of a deterministic CM. We assume that the structure of $F^{(0)}(\boldsymbol{x})$ and $F_i(\boldsymbol{x})$ is such that the CM is one-dimensional (there are no further degenerate ecosystem variables) and that it is the only stable state in the interior region $x_i > 0$. A separation of timescales is present if the system collapses to the CM much faster than the stochastic dynamics. In practical terms, the timescale of collapse can be inferred as the inverse of the non-zero eigenvalues of the system, linearised about the CM [34], while the timescale of fluctuations will be of order R^{-2} [35]. When $\varepsilon > 0$, the timescale elimination procedure can still be applied if $\varepsilon \approx \mathcal{O}\left(R^{-2}\right)$. The effective one-dimensional description of the system now takes the form

$$\dot{x_1} = -\varepsilon\,\mathcal{D}(x_1) + R^{-2}\,\mathcal{S}(x_1) + R^{-1}\zeta(t)\,, \tag{17}$$

where the term $\mathcal{D}(x_1)$ is the deterministic contribution to the effective dynamics and $\mathcal{S}(x_1)$ is the stochastic contribution, while $\zeta(t)$ is an effective noise term. The form these functions take is dependent on $F^{(0)}(\boldsymbol{x},\boldsymbol{e})$, $F^{(\varepsilon)}(\boldsymbol{x},\boldsymbol{e})$ and $F_i(\boldsymbol{x},\boldsymbol{e})$, as well as the noise correlation structure, $H^{(0)}(\boldsymbol{x},\boldsymbol{e})$; however it is independent of the structure of the demographic noise acting on the ecosystem variables (see Eqs. (E30), (E19) and (E20)).

The core assumption we have made to derive Eq. (17) is essentially that the system's ecological processes act on a faster timescale than its evolutionary processes. Even in this general setting, insights about the system's stochastic dynamics can still be drawn (see Appendix E). If $\varepsilon = 0$, the fixation probability of phenotype X_1 is independent of the initial conditions of the ecosystem variables $\boldsymbol{e}$. In fact it is equal to the initial fraction of X_1 in the population, $n_{10}/(n_{10}+n_{20})$. The invasion probability of mutant X_1 phenotype fixating in a resident X_2 population however depends on the stationary state of the X_2 population; this defines the initial invasion conditions (the denominator for the fixation probability of X_1). Denoting by N_1 and N_2 the average numbers of phenotypes X_1 and X_2 in their respective stationary states, we find $\phi_1 = 1/N_2$ and $\phi_2 = 1/N_1$, generalizing Eq. (7). Therefore, for $\varepsilon = 0$ the phenotype that exists at higher densities is more likely to invade and fixate than its competitor, a consequence of its robustness to invasions. This result holds for any choice of finite R. In an ensemble of disconnected populations subject to repeated invasions, we would observe the emergence of high density phenotypes if this phenotype does not carry a cost. While this seems like a reasonable and indeed natural conclusion, it is one entirely absent from the deterministic analysis.

If $\varepsilon > 0$, general results for the phenotype fixation probabilities cannot be obtained. However, if $N_1 > N_2$, in the limit $\varepsilon \to 0$ we have shown that $\phi_1 > \phi_2$. From this, it can be inferred that the term $\mathcal{S}(x_1)$ is positive on average along the slow manifold (see Eq. (E26)). Therefore, if phenotype X_1 exists at higher densities in isolation than phenotype X_2, there will exist a stochastically-induced pressure favoring the invasion of phenotype X_1. Meanwhile, by construction we expect the form of $\mathcal{D}(x_1)$ to be positive, since phenotype X_1 exploits the ecosystem environment less effectively than phenotype X_2. There is therefore a trade-off for competing phenotypes between increasing their phenotype population density and increasing their per capita growth rate. Note that the noise-induced selection function $\mathcal{S}(x_1)$ need not be strictly positive; indeed it may become negative along regions of the SM. This potentially allows for stochastically induced 'fixed points' along the SM, around which the system might remain for unusually large periods of time. This may provide a theoretical understanding of the coexistence behavior observed in [36].

The term $\mathcal{S}(x_1)$ is moderated by factor R^{-2} (see Eq. (8)), or more physically, the typical size of the population. The stochastically induced selection for the high-density phenotype therefore becomes weaker as typical system sizes increase. The trade-off will be most crucial in small populations, or as illustrated in the public good model, systems with a spatial component. If the phenotypes and ecosystem variables move sufficiently slowly in space, the results of Eqs. (13) and (14) can be imported, with the understanding that ϕ_1 and ϕ_2 must be calculated for the new model under consideration.

It is worth noting that the precise functional form of ϕ_1 and ϕ_2 identified in the deterministically neutral case ($\varepsilon = 0$) is dependent on the assumption that phenotype noise fluctuations are uncorrelated. While cor-

related fluctuations (for instance resulting from mutual predation of the phenotypes) can still be addressed with similar methods to those employed here, there is then the potential for the emergence of further noise-induced selection terms (see Appendix F). Careful specification of the phenotype interaction terms is therefore needed to determine to what degree these additional processes might amplify or dampen the induced selection we have identified.

IV. DISCUSSION

In this paper, we have shown that stochastic effects can profoundly alter the dynamics of systems of phenotypes that change the carrying capacity of the total population. Most strikingly, selection can act in the opposite direction from that of the deterministic prediction if the phenotype that is deterministically selected for also reduces the carrying capacity of the population. The methods used to analyze the models outlined in the paper are based on the removal of fast degrees of freedom [30]. The conclusions drawn are therefore expected to remain valid as long as the rate of change of the phenotype population composition occurs on a shorter timescale than the remaining ecological processes.

By illustrating this phenomenon in the context of public good production, we have revealed a mechanism by which the dilemma of cooperation can be averted in a very natural way: by removing the unrealistic assumptions of fixed population size inherent in Moran-type game theoretic models. The potential for such behavior has been previously illustrated with the aid of a modified Moran model [17] and a single variable Wright-Fisher type model [18] that assumes discrete generations. However we have shown that the mechanism can manifest more generally in multivariate continuous time systems. Our analysis may also provide a mathematical insight into the related phenomenon of fluctuation-induced coexistence that has been observed in simulations of a similar public good model featuring exogenous additive noise [36]: such coexistence may rely on a similar conflict between noise-induced selection for producing phenotypes and deterministic selection against them.

For biologically reasonable public good production costs, selection reversal is only observed in systems that consist of a very small number of individuals. However, by building a metapopulation analogue of the model to account for spatial structure, the range of parameters over which selection reversal is observed can be dramatically increased, so long as public good diffusion and phenotype dispersal between populations are not large. Two distinct mechanisms are responsible for these results. First, including spatial structure allows for small, local effective population sizes, even as the total size of the population increases. This facilitates the stochastic effects that lead to selection reversal. Second, since producer populations tend to exist at greater numbers (or higher local densities) they produce more migrants. The stochastic advantage received by producers is thus amplified, as not only are they more robust stochastically to invasions, but also more likely to produce invaders. Away from the low-dispersal, zero public good diffusion limit, the effect of selection reversal is diminished, but is still present across a range of biologically reasonable parameters. The analytical framework we have outlined may prove insightful for understanding the simulation results observed in [37], where a similar metapopulation public good model was considered. In addition to fixation of producers (in the low dispersal-diffusion limit) and fixation of non-producers (in the high dispersal-diffusion limit), [37] observed an intermediate parameter range in which noise induced coexistence was possible. Though our model does not feature such a regime, extending our mathematical analysis to their model would be an interesting area for future investigations. However it must be noted that coexistence in a stochastic setting is inherently difficult to quantify analytically, as for infinite times some phenotype will always go extinct.

That space can aid the maintenance of cooperation is well known [38, 39]. Generally, however, this is a result of spatial correlations between related phenotypes, so that cooperators are likely to be born neighboring other cooperators (and share the benefits of cooperation) while defectors can only extract benefits at the perimeter of a cooperating cluster. This is not what occurs in the model presented in this paper. Indeed, while we have assumed in our analytic derivation of the invasion probability that dispersal is small enough that each patch essentially contains a single phenotype, we find that the phenomenon of selection reversal manifests outside this limit (see Appendix H, movie S2 in which a majority of patches contain a mix of producers and non-producers). Instead, producing phenotypes have a selective advantage due to the correlation between the fraction of producers on a patch and the total number of individuals on a patch, which provides both resistance to invasions and an increased dispersal rate.

Most commonly in spatial game theoretic models of cooperation-defection, individuals are placed at discrete locations on a graph [40, 41]. In contrast, by using a metapopulation modelling framework we have been able to capture the effect of local variations in phenotype densities across space, which is the driver of selection amplification in our model. Nevertheless, the question that remains is which modelling methodology is more biologically reasonable. This clearly depends on the biological situation. However, in terms of test-ability, our model makes certain distinct predictions. In [41], producers and non-producers were modeled as residing on nodes of a spatial network, with a public good diffusing between them. The investigation concludes that both lower public good diffusion and lower spatial dimensions (e.g. systems on a surface rather than in a volume) should encourage public good production, essentially by limiting the 'surface area' of producing clusters. While our inves-

tigation certainly predicts that lower public good diffusion is preferable, stochastic selection reversal does not require that the spatial dimension of the system is low. In fact the result utilized in Eq. (12) holds for patches arranged on any regular graph (where each vertex has the same number of neighbors), and thus could be used to describe patches arranged on a cubic, or even hexagonal, lattice.

In our final investigation, we have shown that stochastic selection reversal is not an artifact of a specific model choice, but may be expected across a wide range of models. These models consist of two phenotypes, competing under weak deterministic selection strength, reproducing according to replicator dynamics and interacting with their environment. Thus the phenomenon of selection reversal is very general; however, it depends strongly on how one specifies a selective gradient. We take one phenotype to have a stochastic selective advantage over the other if a single mutant is more likely to invade a resident population of the opposite type. Such a definition is also used in standard stochastic game theoretic models [4]. A key difference here however (where the population size is not fixed) is that the invasion probability is not specified by a unique initial condition; we must also specify the size of the resident population. We have assumed that the invading mutant encounters a resident population in its stationary state. This is by no means an unusual assumption; it is the natural analogue of the initial conditions in a fixed population size model. Essentially it assumes a very large time between invasion or mutation events, an approach often taken in adaptive dynamics [42].

If instead we assumed a well-mixed system far from the steady state, our results would differ. For instance, suppose the system initially contains equal numbers of the two phenotypes. For the case when the two phenotypes have equal reproductive rates ($\varepsilon = 0$), the phenotypes have equal fixation probability. For $\varepsilon > 0$, the phenotype with the higher birth rate has the larger fixation probability, regardless of its influence on the system's carrying capacity. This apparent contradiction with the results we developed in the body of the paper echos the observations of $r - K$ selection theory [43]: selection for higher birth rates (r-selection) acts on frequently disturbed systems that lie far from equilibrium, while selection for improved competitive interactions or carrying capacities (K-selection) acts on rarely disturbed systems. In addition, $r - K$ selection theory suggests that K-selected species are typically larger in size and, as a consequence, consist of a lower number of individuals [19]. This indicates a further parallel with our stochastic model framework, since selection for higher carrying capacities requires that the typical number of individuals (of both the low and high carrying capacity phenotypes) is small. Though the mechanism that leads us to these conclusions is distinct, our stochastic analysis provides a complementary view of $r - K$-selection theory, which may be applicable to simple microorganisms. In exploring this analogous behavior further, future investigations may also benefit from considering the results of [15], where it was shown that stochastically induced selection can change direction near carrying capacity.

Although we have implicitly developed our results in the low mutation limit, including mutation explicitly in the modeling framework is possible. This would be an interesting extension to the framework. In the well-mixed scenario, it is likely that the inclusion of mutation will complicate the intuition developed here: while larger populations are more robust to invasions, they are also more prone to mutations, by virtue of their size. While this may be offset by the additional benefits garnered in the spatial analogue of the model, a complex set of timescale-dependent behaviors is likely to emerge.

Finally, we propose a rigorous analytical investigation of existing models that conform to the framework we have outlined; an example is the work conducted in [36, 37], which we believe to be mathematically explainable within our formalism. In the context of induced selection, whereby deterministically neutral systems become non-neutral in the stochastic setting, similar ideas have already been extended to disease dynamics [16] and the evolution of dispersal [44, 45]. The extension of selection reversal to such novel ecological models may provide further insight. Furthermore, this general scheme may be of relevance to many other systems in ecological and biological modeling, such as cancer, for which the evolution of phenotypes that profoundly alter cell carrying capacity can be of primary importance.

ACKNOWLEDGMENTS

TR acknowledges funding from the Royal Society of London.

Appendix A: Obtaining the SDE system from the microscopic individual based model

We begin with a model consisting of a discrete number of entities, two phenotypes of a species, X and Y and a public good Q. They interact according to the transitions

$$\begin{aligned} X \underset{\kappa/R^2}{\overset{b_x}{\rightleftarrows}} X + X\,, \quad Y + X \xrightarrow{\kappa/R^2} X\,, \\ Y \underset{\kappa/R^2}{\overset{b_y}{\rightleftarrows}} Y + Y\,, \quad Y + X \xrightarrow{\kappa/R^2} Y\,, \\ X + Q \xrightarrow{r/R^2} X + X + Q\,, \quad Y + Q \xrightarrow{r/R^2} Y + Y + Q\,, \\ X \xrightarrow{p_x} X + Q\,, \quad Q \xrightarrow{\delta} \varnothing\,. \end{aligned} \tag{A1}$$

The term R^{-2} occurs in all terms involving two reactants. It thus controls the interaction probability between instances of the phenotypes and the public good. Taking larger R decreases the interaction probability of phenotypes X and Y and the public good and allowing the

populations to grow to greater numerical abundances. The parameter R can thus be understood as a measure of the spatial scale of the system; when R is increases, the probability of interactions in the well-mixed system is decreased while the number of individuals the system can contain is increased.

Let us denote $\boldsymbol{n} = (n_x,\ n_y,\ n_q)$ the numbers of X, Y and Q respectively. Then the dynamics of this system can be described by the set of partial difference equations

$$\frac{dP(\boldsymbol{n},t)}{dt} = \sum_{\boldsymbol{n}'\neq\boldsymbol{n}} \left[T(\boldsymbol{n}|\boldsymbol{n}')P(\boldsymbol{n}',t) - T(\boldsymbol{n}'|\boldsymbol{n})P(\boldsymbol{n},t)\right] , \tag{A2}$$

where $P(\boldsymbol{n},t)$ is the probability of the state being in state $\boldsymbol{n}$ at time t, and $T(\boldsymbol{n}'|\boldsymbol{n})$, the probability transition rate, is the probability per unit time of transitioning from state $\boldsymbol{n}$ to $\boldsymbol{n}'$. Formally this is known as the master equation [47]. Given the reactions Eq. (A1) the probability transition rates can be expressed as

$$\begin{aligned}
T_1(n_x+1,n_y,n_q|n_x,n_y,n_q) &= b_x n_x + \frac{r}{R^2} n_x n_q \,, \\
T_2(n_x,n_y+1,n_q|n_x,n_y,n_q) &= b_y n_y + \tfrac{r}{R^2} n_y n_q \,, \\
T_3(n_x-1,n_y,n_q|n_x,n_y,n_q) &= \frac{\kappa}{R^2} n_x \left(n_x+n_y\right) \,, \\
T_4(n_x,n_y-1,n_q|n_x,n_y,n_q) &= \tfrac{\kappa}{R^2} n_y \left(n_x+n_y\right) \,, \\
T_5(n_x,n_y,n_q+1|n_x,n_y,n_q) &= p_x n_x \,, \\
T_6(n_x,n_y,n_q-1|n_x,n_y,n_q) &= \delta n_q \,.
\end{aligned} \tag{A3}$$

Let us now make a change of variables into the scaled expressions $\boldsymbol{x} = (x,y,q) = (n_x,n_y,n_q)/R^2$. Substituting the probability transition rates into Eq. (A2), we find recurrent factors of $1/R^2$ appearing in the resulting expression. These terms are associated with the local transitions from state $\boldsymbol{n}$ to the surrounding states. If R^2 is sufficiently large, the population grows larger (as the crowding terms in Eq. (A1) grow small). We may then Taylor expand Eq. (A2) in R^{-1}, assuming that the variables (x,y,q) are approximately continuous [28]. Truncating at second order in R^{-4}, we arrive at a partial differential equation for $p(x,y,q,t)$ of the form

$$\begin{aligned}
\frac{\partial p(\boldsymbol{x},t)}{\partial t} = &- \frac{1}{R^2}\sum_i \frac{\partial}{\partial x_i}\left[A_i(\boldsymbol{x})p(\boldsymbol{x},t)\right] \\
&+ \frac{1}{2R^4}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}\left[B_{ij}(\boldsymbol{x})p(\boldsymbol{x},t)\right] \,, \\
\boldsymbol{x} &= (x_1,x_2,x_3) \equiv (x,y,q) \,.
\end{aligned} \tag{A4}$$

This is a diffusion approximation in a population genetics context [22], but more generally is akin to the Kramers-Moyal expansion [28] or a nonlinear analogue of the van Kampen expansion [47]. The forms of $\boldsymbol{A}(\boldsymbol{x})$ and $B(\boldsymbol{x})$, given transition rates Eq. (A3) are found to be

$$\begin{aligned}
A_x(\boldsymbol{x}) &= x\left(b_x + rq - \kappa(x+y)\right) \,, \\
A_y(\boldsymbol{x}) &= y\left(b_y + rq - \kappa(x+y)\right) \,, \\
A_q(\boldsymbol{x}) &= p_x x - \delta q \,,
\end{aligned} \tag{A5}$$

and

$$\begin{aligned}
B_{xx}(\boldsymbol{x}) &= x\left(b_x + rq + \kappa x + \kappa y\right) \,, \\
B_{yy}(\boldsymbol{x}) &= y\left(b_y + rq + \kappa x + \kappa y\right) \,, \\
B_{qq}(\boldsymbol{x}) &= p_x x + \delta q \,, \\
B_{ij} &= 0 \quad \forall \quad i \neq j \,.
\end{aligned} \tag{A6}$$

Further, it can be shown that the above PDE is equivalent to the set of Itō SDEs [48]

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\tau} = \boldsymbol{A}(\boldsymbol{x}) + \frac{1}{R}\boldsymbol{\eta}(\tau) \,, \tag{A7}$$

where $\tau = tR^2$ and $\boldsymbol{\eta}(t)$ are Gaussian white noise terms with zero mean and correlations

$$\langle \eta_i(\tau)\eta_j(\tau')\rangle = \delta(\tau-\tau')B_{ij}(\boldsymbol{x}) \,. \tag{A8}$$

Notice that the correlations are multiplicative and thus dependent on the state of the system.

Appendix B: Obtaining one-dimensional effective public good model

In this section we seek to identify and remove the fast-modes of the SDE system Eq. (A7), and thus obtain an effective one-dimensional description of the dynamics. We make use of methods of fast-mode elimination described in [30]. Firstly we note that the deterministic nullcline for q is given by

$$q = \frac{p_x x}{\delta} \equiv Z_q(x,y) \,. \tag{B1}$$

Therefore, if the production and decay of public good occur much faster than the processes associated with the phenotypes, we would expect the public good to quickly attain this value, after which its dynamics would be slaved to those of x and y. Notice that deterministically, substituting Eq. (B1) into Eq. (A5) recovers a Lotka-Volterra competition model for two competing species.

To make further analytic progress, we begin by considering the quasi-neutral limit in which $b_x = b_y \equiv b$. Under these conditions, the deterministic system exhibits a center manifold (CM) given by Eq. (B1) and

$$y = \frac{\left[b\delta - (\delta\kappa - rp_x)x\right]}{\delta\kappa} \equiv Z_y(x) \,. \tag{B2}$$

The CM is stable for $\kappa\delta > rp_x$, and we assume that this condition holds throughout the paper. Calculating the intersection of the center manifold at the boundaries $y = 0$ and $x = 0$ allows us to determine the mean population size in the quasi-neutral ($\varepsilon = 0$) limit when it consists of only producers and non-producers respectively;

$$N_x^{(0)} = R^2 K_x^{(0)} \,, \quad K_x^{(0)} = \left(\frac{b\delta}{\delta\kappa - rp_x}\right) \,, \tag{B3}$$

$$N_y^{(0)} = R^2 K_y^{(0)} \,, \quad K_y^{(0)} = \left(\frac{b\delta}{\delta\kappa - rp_y}\right) \,. \tag{B4}$$

These parameters will be useful in the following analysis.

Deterministically, the system comes to rest on a point along the CM (defined by Eqs. (B1) and(B2)), which depends on the system's initial conditions. When stochasticity is included, the CM ceases to exist in any true sense. However, when the noise is small (already assumed in the derivation of SDEs (A7)) we can say that far from the CM, we expect the dynamics to be dominated by the deterministic collapse to the CM, while in the vicinity of the CM, we expect noise to play a more important role, driving the slow change in population composition until one or other of the phenotypes fixates. We wish to exploit this timescale separation, and obtain an effective description of the dynamics in terms of a single variable.

To begin, we note that the stochastic dynamics along the CM has two components. First, noise can move the system neutrally *along* the CM. Second, noise can take the system *off* the CM, at which point we expect the deterministic component of the dynamics to become more prevalent, driving the system back to the CM. In order to capture the effect of both of these processes on the effective dynamics along the CM, we implement a non-linear projection of the stochastic system to the CM. Essentially this assumes that fluctuations which take the system away from the manifold are instantaneously mapped along deterministic trajectories back to the CM. In order to formalize this, the mapping $z = f(x, y, q)$ is introduced, where $f(x, Z_y(x), Z_q(x)) = x$; that is z gives the position on the CM, parameterized by x, which intersects a deterministic trajectory beginning at (x, y, q). The mapping can be determined analytically from the observation that the quantity x/y in Eq. (A7) is invariant in this quasi-neutral ($b_x = b_y$) scenario. Therefore

$$\frac{z}{Z_y(z)} = \frac{x}{y}\,, \quad z = \frac{b\delta x}{(\delta\kappa - pr)x + \delta\kappa y}\,. \tag{B5}$$

The effective dynamics for z can now be straightforwardly calculated by differentiating Eq. (B5) with respect to t. One must note however that since the original SDE system is defined in the Itō sense, the normal rules of calculus no longer apply. Applying Itō's rules of calculus appropriately [30, 47], we find that the effective dynamics along the CM take the following form

$$\dot{z} = \frac{1}{R^2}\mathcal{S}(z) + \frac{1}{R}\zeta(t)\,, \tag{B6}$$

where

$$\begin{aligned}\mathcal{S}(z) &= \frac{1}{2}\left(\frac{\partial^2 z}{\partial x^2}B_{xx}(\boldsymbol{x}) + \frac{\partial^2 z}{\partial y^2}B_{yy}(\boldsymbol{x})\right)|_{x=z,y=Z_y(z),q=Z_q(z)}\,, \\ &= \frac{2p_x r}{\delta}z\left\{1 + \frac{1}{b^2\delta^2}z\left[b\delta\left(2p_x r - \delta\kappa\right) + p_x r\left(p_x r - \delta\kappa\right)z\right]\right\} \\ &= 2b\left(\frac{K_x^{(0)} - K_y^{(0)}}{(K_x^{(0)})^3(K_y^{(0)})^2}\right)z\left(K_x^{(0)} - z\right)\left[K_x^{(0)}K_y^{(0)}\right. \\ &+ \left.\left(K_x^{(0)} - K_y^{(0)}\right)z\right]\,,\end{aligned} \tag{B7}$$

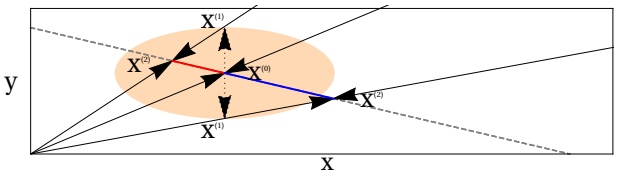


FIG. 6. Figure illustrating the origin of stochastically induced drift along the center manifold (CM). The gray dashed line shows the form of the deterministic center manifold, which intersects the x axis at a higher value than the y axis (phenotype X has a higher carrying capacity due to the production of the public good). The red shaded circle illustrates the form of the Gaussian noise centered on the point $\boldsymbol{x}^{(0)}$ on the CM. Fluctuations in the population are equally likely to increase or decrease the frequency of the Y phenotype to the points $\boldsymbol{x}^{(1)}$. Away from the CM, the deterministic pressure to the CM becomes prominent, forcing the system along quasi-deterministic trajectories back to the CM, at the points $\boldsymbol{x}^{(2)}$. The resulting distribution of $\boldsymbol{x}^{(2)}$ does not have a mean centered on $\boldsymbol{x}^{(0)}$. Rather, the distribution is shifted, inducing a drift in favor of the producing X phenotype.

and

$$\langle\zeta(t)\rangle = 0\,, \quad \langle\zeta(t)\zeta(t')\rangle = \delta(t-t')\mathcal{B}(z)\,,$$

with

$$\begin{aligned}\mathcal{B}(z) &= \left(\left[\frac{\partial z}{\partial x}\right]^2 B_{xx}(\boldsymbol{x}) + \left[\frac{\partial z}{\partial y}\right]^2 B_{yy}(\boldsymbol{x})\right)|_{x=z,y=Z_y(z),q=Z_q(z)}\,, \\ &= 2z\left\{b + \frac{1}{b^2\delta^3}z\left[b^2\delta^2\left(3p_x r - \delta\kappa\right)\right.\right. \\ &+ \left.\left. bp_x r\delta\left(3p_x r - 2\delta\kappa\right)z p_x^2 r^2\left(p_x r - \delta\kappa\right)z^2\right]\right\}\,, \\ &= 2b\left(\frac{1}{(K_x^{(0)})^3(K_y^{(0)})^2}\right)z\left(K_x^{(0)} - z\right)\left[K_x^{(0)}K_y^{(0)}\right. \\ &+ \left.\left(K_x^{(0)} - K_y^{(0)}\right)z\right]^2\,.\end{aligned} \tag{B9, B10}$$

Notice that since the mapping Eq. (B5) is independent of q, both Eq. (B7) and Eq. (B9) do not depend on the noise correlations in q.

While the deterministic system features no dynamics along the CM, the effective SDE (B6) does feature a drift in the mean state, embodied by $\mathcal{S}(z)$. Understanding the origin of this induced drift term requires considering the following. We envisage fluctuations arising from a single point on the CM, $\boldsymbol{x}^{(0)}$, which take to the system to point off the CM, $\boldsymbol{x}^{(1)}$ (see Fig. 6). The point $\boldsymbol{x}^{(1)}$ is clearly stochastic, but its distribution is approximately Gaussian, with a variance defined by $B(\boldsymbol{x}^{(0)})$. The fluctuation is now mapped back along a deterministic trajectory to a point $\boldsymbol{x}^{(2)}$ on the CM. The location $\boldsymbol{x}^{(2)}$ is also stochastic (dependent as it is on $\boldsymbol{x}^{(1)}$), and has its own distribution. The presence of the term $\mathcal{S}(z)$ in Eq. (B6) is indicative of the fact that the mean of the distribution of $\boldsymbol{x}^{(2)}$ is not $\boldsymbol{x}^{(0)}$; fluctuation events on average are mapped

back to the CM with a preferred direction, inducing drift along the CM. Note that $\mathcal{S}(z)$ is positive along the length of the CM, which is defined on the interval $[0, K_x^{(0)}]$.

We now turn our attention to the case when $\varepsilon > 0$. So long as ε is small, a separation of timescales is still present, though now no center manifold exists. Instead there is a slow manifold (SM), to which the deterministic system quickly relaxes, before slowly moving along it until phenotype Y fixates. The equations for the population size at the boundaries of the SM are formally given by

$$\begin{aligned} N_x &= R^2 K_x\,, \quad K_x = \left(\frac{b_x\delta}{\delta\kappa - r p_x}\right) \equiv K_x^{(0)} + \mathcal{O}(\varepsilon)\,, \\ N_y &= R^2 K_y\,, \quad K_y = \left(\frac{b_y\delta}{\delta\kappa - r p_y}\right) \equiv K_x^{(0)} + \mathcal{O}(\varepsilon)\,. \end{aligned} \tag{B11}$$

In order to proceed with the stochastic calculation, we assume $\varepsilon \approx \mathcal{O}(R^{-2})$, and work order by order in R^{-1}. At leading order, the equation for the SM is identical to that of the CM, Eqs. (B1) and (B2). The mapping to the SM is also unchanged at leading order from the quasi-neutral case (see Eq. (B5)). We proceed as before to obtain an effective description of the system dynamics in terms of z [30], now obtaining the dynamics,

$$\dot{z} = -\varepsilon\mathcal{D}(z) + \frac{1}{R^2}\mathcal{S}(z) + \frac{1}{R}\zeta(t)\,. \tag{B12}$$

where

$$\begin{aligned} \mathcal{D}(z) &= -\left(\frac{dz}{dx}A_x(\boldsymbol{x}) + \frac{dz}{dy}A_y(\boldsymbol{x})\right)|_{x=z,y=Z_y(z),q=Z_q(z)}\,, \\ &= bz\left[1 - \left(\frac{\delta\kappa - p_x r}{b\delta}\right)z\right]\,, \\ &= \frac{b}{K_x^{(0)}} z\left(K_x^{(0)} - z\right)\,, \end{aligned} \tag{B13}$$

and $\mathcal{S}(t)$ and $\zeta(t)$ retain their form from the quasi-neutral case, Eqs. (B8) and (B10). The function $\mathcal{D}(z)$ is the deterministic contribution to the dynamics along the SM. This expression is that which would be obtained using standard fast variable elimination techniques on the deterministic system. From Eq. (B13), we can see that $\mathcal{D}(z)$ is positive along the length of the SM and therefore acts (as we would expect) to increase the selective advantage of the non-producers, phenotype Y. There is therefore a conflict between the two components of the drift in the system. The term $\mathcal{D}(z)$ works against producers along the length of the SM, while $\mathcal{S}(z)$ creates a selective pressure in favor of producers. Ultimately, which term is more prevalent is dependent on the parameters ε and R (see Eq. (B12)); small R leads to a small population size in which stochastic effects are stronger, and so producers are more likely to be selected for. In contrast, when the deterministic cost for good production is increased, the non-producers have an increased advantage over producers.

Adopting the notation used in the main text, in which we set $z = x$ (which is valid on the CM and SM at leading order), the expression for the SDE (B12) can alternatively be written

$$\dot{x} = \frac{b}{K_x^{(0)}} x\left(K_x^{(0)} - x\right)\left(\frac{1}{R^2}\,\mathcal{F}(x) - \varepsilon\right) + \frac{1}{R}\zeta(t)\,, \tag{B14}$$

where

$$\mathcal{F}(x) = 2\left(\frac{K_x^{(0)} - K_y^{(0)}}{(K_x^{(0)})^2 (K_y^{(0)})^2}\right)\left[K_x^{(0)} K_y^{(0)} + \left(K_x^{(0)} - K_y^{(0)}\right)x\right]\,. \tag{B15}$$

Appendix C: Probability of fixation for the reduced public good model

The fixation probability for a phenotype in a single variable system can be calculated using standard methods [28]. In order to conduct the calculation, we need expressions for the absorbing boundaries of the problem. For the reduced system given in Eq. (B12), these lie at $z = 0$ and $z = K_x^{(0)}$. The fact that the boundary for the problem exists at $z = K_x^{(0)}$, rather than $z = K_x$, is a consequence of the order to which we are working in ε. At this order the SM is approximated by the expression for the CM, which intersects the absorbing boundaries $x = 0$ and $y = 0$ at $z = 0$ and $z = K_x^{(0)}$ respectively. Denoting $Q(z_0)$ the fixation probability of producing phenotype X given an initial frequency z_0 on the CM/SM, the fixation probability can be conveniently be expressed

$$\begin{aligned} Q(z_0) &= \frac{\int_{z=0}^{z_0}\psi(z)dz}{\int_{z=0}^{K_x}\psi(z)dz}\,, \\ \psi(z) &= \exp\left[\int_0^z \frac{2(-\varepsilon R\mathcal{D}(z') + \mathcal{S}(z'))}{\mathcal{B}(z')}dz'\right]\,. \end{aligned} \tag{C1}$$

Substituting for $\mathcal{D}(z)$, $\mathcal{S}(z)$ and $\mathcal{B}(z)$ from Eqs. (B13), (B8) and (B10), we find

$$\begin{aligned} Q(z_0) &= \frac{1 - G(z_0)}{1 - G(K_x)}\,, \\ G(z_0) &= \exp\left[\frac{\left(\varepsilon N_y^{(0)} K_x^{(0)} z_0\right)}{\left(K_x^{(0)} K_y^{(0)} + (K_x^{(0)} - K_y^{(0)})z\right)}\right]\,. \end{aligned} \tag{C2}$$

The nature of these expressions can be understood more intuitively if we move from considering the initial frequency of X on the CM, $z_0 = n_{x0}/R^2$, to considering the initial fraction of phenotype X on the CM, f_{z0}. The fraction and number of phenotype X on the CM are related by

$$\begin{aligned} f_{z0} &= \frac{z}{z + Z_y(z)}\,, \\ z &= \frac{K_x^{(0)} K_y^{(0)} f_{z0}}{K_x^{(0)} - (K_x^{(0)} - K_y^{(0)})f_{z0}}\,. \end{aligned} \tag{C3}$$

Substituting this into Eq. (C2), we find

$$Q(f_{z0}) = \frac{1-\exp\left[\varepsilon N_y^{(0)} f_{z0}\right]}{1-\exp\left[\varepsilon N_y^{(0)}\right]}\,, \qquad Q(f_{z0})|_{\varepsilon=0} = f_{z0}\,. \tag{C4}$$

On first appraisal, the fixation probabilities Eq. (C4) appear to share the form of the well-mixed Moran model with weak selection. There is however one crucial distinction; the relation between f_{z0} and (x_0, y_0, q_0) is dependent on the form of the CM/SM, and is not necessarily symmetric under the interchange of X and Y. For instance, let us consider the quasi-neutral case ($\varepsilon = 0$) with the population initially consisting of a mutant X in a population of the Y phenotype in its stationary state. Then $f_{z0} = 1/N_y$. In contrast, if the mutant is of phenotype Y, and the resident population consists of phenotype X in the stationary state, $f_{z0} = 1 - 1/N_x$. Since N_x and N_y are distinct, these frequencies are not the same, and Eq. (C4) is not symmetric under the interchange of phenotypes, undermining its apparent similarities with the Moran model.

In this section a crucial aspect of the selection reversal has been elucidated. The selection reversal along the SM is a result of the differing densities at which the populations of X and Y phenotypes reside in isolation. In a deterministic system, we would define the fitter phenotype as the one which fixates at long times. In stochastic Moran-type model, the fitter phenotype is defined as that with the greater invasion probability. Since Moran-type models feature a constant population size, N, the invasion probability of a mutant phenotype is defined by a unique initial condition; a single mutant, and $N-1$ residents. In systems such as the public good model discussed in this paper, the invasion probability is no longer defined uniquely by the specification of a single invading mutant; we must also define the size of the resident phenotype population and the public good density. If the system has been allowed to relax to a stationary state before the mutant is introduced, then selection reversal along the CM may be present, and it is possible for the producing phenotype to have a larger fixation probability than the non-producing phenotype. Thus the producing phenotype may be fitter.

Appendix D: Pairwise invasibility for non-producers, producers and hyper-producers

In this section we explore the pairwise invasibility of three separate phenotypes, non-producers, producers and hyper-producers. We begin by noting that, under the assumption that the birth rates differ by only a small amount from phenotype to phenotype, the invasion probability of phenotype i in a resident population j, $\phi_{i|j}$, can be expressed

$$\phi_{i|j} = \frac{1-\exp\left[(b_i-b_j)/(\kappa N_j^{(0)})\right]}{1-\exp\left[(b_i-b_j)R^2/\kappa\right]}\,. \tag{D1}$$

We therefore define phenotype i as fitter than phenotype j if $\phi_{i|j} > \phi_{j|i}$. Let us now explicitly express the birth rates of each of the phenotypes as

$$\begin{aligned} \text{Non}-\text{producer}:\ b_y &= b\,,\\ \text{Producer}:\ b_x &= b(1-\varepsilon)\,,\\ \text{Hyper}-\text{producer}:\ b_u &= b(1-a_b\varepsilon)\,. \end{aligned} \tag{D2}$$

We now wish to obtain an expression for the critical costs to birth rate ε at which producers are fitter than non-producers, hyper-producers are fitter than non-producers and hyper-producers are fitter than producers. To do this we must solve $\phi_{i|j} = \phi_{j|i}$ for ε for each pair of phenotypes. An analytic solution is available if we set $\varepsilon = \tilde{\varepsilon}R^{-2}$ with $\tilde{\varepsilon}$ of order one, and expand Taylor expand in R^{-2}. Truncating at first order, we find that the critical cost for species i to be fitter than species j, $\varepsilon_{i|j}$ is given by

$$\varepsilon_{i|j} = \frac{\kappa \log\left[(p_i r - \delta\kappa)/(p_j r - \delta\kappa)\right]}{\left[(b_i-b_j)/\varepsilon\right] R^2}\,. \tag{D3}$$

We note that this provides eight different possible scenarios of fitness ranking, described in Fig. 7. Substituting in our equations for the birth rates, Eq. (D2), these expressions become

$$\varepsilon_{x|y} = \frac{\kappa}{bR^2}\log\left[-\frac{\delta\kappa}{p_x r - \delta\kappa}\right]\,, \tag{D4}$$

$$\varepsilon_{u|y} = \frac{\kappa}{abR^2}\log\left[-\frac{\delta\kappa}{p_u r - \delta\kappa}\right]\,, \tag{D5}$$

$$\varepsilon_{u|x} = \frac{\kappa}{(a-1)bR^2}\log\left[\frac{p_x r - \delta\kappa}{p_u r - \delta\kappa}\right]\,. \tag{D6}$$

Clearly the exact scenarios which emerge for a given set of parameters depends on the relationship between p_x and p_u. We make the assumption

$$p_u = a_p p_x\,. \tag{D7}$$

For $a_p > a_b$, the hyper-producer pays a discounted cost to its birth rate for its additional good production. In this situation, only scenarios (c-f) are possible in Fig. 7. It is always better to be a hyper-producer or a non-producer, depending on the production rate p_x and ϵ. This 'all or nothing' result makes intuitive sense; if the hyper-producer produces much more than the producer, but pays only fractionally more to its birth rate, any region in which production is favored will be disproportionately advantageous to the hyper-producers. In contrast, if $a_p < a_b$, the hyper-producer receives decreasing production returns as a function of the cost it pays to birth in comparison with the producer. In this case, scenarios (a-b) and (e-f) are possible. Either producers or non-producers are favored, and hyper-producers are never favored.

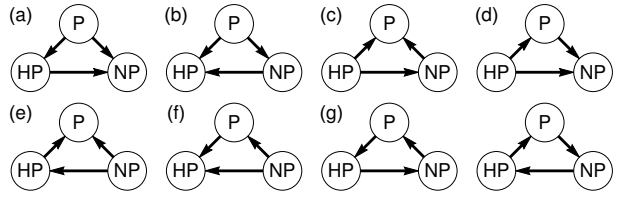


FIG. 7. Eight different fitness rankings are possible based on the pairwise invasibility probabilities of non-producers, producers and hyper-producers. (A) Producers have a larger invasion probability than both hyperproducers and nonproducers, while hyperproducers have a larger invasion probability than nonproducers. (B) Producers have a larger invasion probability than both hyperproducers and nonproducers, while nonproducers have a larger invasion probability than hyperproducers. (C) Hyperproducers have a larger invasion probability than both producers and nonproducers, while nonproducers have a larger invasion probability than producers. (D) Hyperproducers have a larger invasion probability than both producers and nonproducers, while producers have a larger invasion probability than producers. (E) Nonproducers have a larger invasion probability than both producers and hyperproducers, while hyperproducers have a larger invasion probability than producers. (F) Nonproducers have a larger invasion probability than both producers and hyperproducers, while producers have a larger invasion probability than hyperproducers. (G) Producers have a larger invasion probability than hyperproducers. Hyperproducers have a larger invasion probability than nonproducers. Nonproducers have a larger invasion probability than producers. (H) Producers have a larger invasion probability than nonproducers. Nonproducers have a larger invasion probability than hyperproducers. Hyperproducers have a larger invasion probability than producers. The nontransitive dynamics of G and H are not seen in the public good model.

Appendix E: Generality of results

We begin by specifying in a very general way the dynamics of an arbitrary IBM with m distinct types of constituent, fully described by a set of u reaction rates. The model can be expressed in chemical reaction notation as

$$\sum_{i=1}^{m} a_{\mu i} X_i \xrightarrow{r_\mu} \sum_{i=1}^{m} b_{\mu i} X_i, \quad \forall \mu = 1, \dots u, \tag{E1}$$

where $a_{\mu i}$ and $b_{\mu i}$ respectively specify the reactants and products of the μ^{th} reaction, and r_μ are the reaction rate constants (see, for example, Eq. (A1)). The stoichiometric matrix is defined by $\nu_{i\mu} = b_{\mu i} - a_{\mu i}$, whose elements give the change in number of the i^{th} species due to the μ^{th} reaction. Together with the rate constants r_μ, the stoichiometric matrix allows us to express the transition rates

$$T_\mu(\boldsymbol{n} + \boldsymbol{\nu}_\mu | \boldsymbol{n}) = r_\mu \prod_{i=1}^{m} a_{\mu i} \frac{n_i}{R^2}, \tag{E2}$$

where R^2 once again is a controlled measure of how often constituents interact (see Eq. (A3)). In the well-mixed model, it therefore directly controls the typical area of the system. Together with the master equation (A2), the full stochastic dynamics are specified.

With a general notation now in hand, we now begin to define the specific type of system we will analyze. We consider a system consisting of two phenotypes, X_1 and X_2, who interact with a set of discrete ecosystem variables X_i, for $i = 3, \dots, N$. The state of the system at any time is given by the number of each phenotype and ecosystem constituent $\boldsymbol{n} = (n_1, n_2, n_3, \dots, n_N)$. The situation we envisage is as follows; while the interplay between the phenotypes and the ecosystem is relevant for the dynamics, we are primarily interested in the evolutionary dynamics and outcome of competition between the two phenotypes. We make the following assumptions on their dynamics;

1. Each phenotype birth and death event is proportional to the number of that phenotype;

$$\begin{aligned} &\text{if} \quad \nu_{1\mu} \neq 0 \quad \text{then} \quad a_{\mu 1} > 0\,, \qquad \text{and} \\ &\text{if} \quad \nu_{2\mu} \neq 0 \quad \text{then} \quad a_{\mu 2} > 0\,. \end{aligned} \tag{E3}$$

2. The phenotypes are very similar in their utilization of the ecosystem. For each μ^{th} reaction that changes the frequency of X_1, there therefore exists a similar reaction μ' that changes the frequency of X_2 such that;

$$\nu_{1\mu} r_\mu = \nu_{2\mu'} (r_{\mu'} + \mathcal{O}(\varepsilon))\,. \tag{E4}$$

3. There is no reaction which simultaneously changes the frequencies of the phenotypes (i.e. no cannibalization or simultaneous killing);

$$\nu_{1\mu} \nu_{2\mu} = 0 \qquad \forall\, \mu\,. \tag{E5}$$

The phenotypes may however differ significantly in their effect on the ecosystem, so that one phenotype may deplete or increase ecosystem constituents in an entirely distinct way to the other (for instance, the production of a public good by phenotype X in Eq. (A1)).

As R is increases so too does the number of each phenotype and ecosystem constituent. If R is sufficiently large, once again a system-size expansion of the master equation can be conducted. Making the change of variables $x_1 = n_1/R^2$, $x_2 = n_2/R^2$ and $e_i = n_{i-2}/R^2$, we obtain the set of Itō SDEs

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= x_1 \left[F^{(0)}(\boldsymbol{x}, \boldsymbol{e}) - \varepsilon F^{(1)}(\boldsymbol{x}, \boldsymbol{e}) \right] + \frac{1}{R} \eta_1(t)\,, \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= x_2 F^{(0)}(\boldsymbol{x}, \boldsymbol{e}) + \frac{1}{R} \eta_2(t)\,, \\ \frac{\mathrm{d}e_i}{\mathrm{d}t} &= h_i(\boldsymbol{x}, \boldsymbol{e}) + \frac{1}{R} \beta_i(t)\,, \quad \forall\, i = 1, \dots N\,. \end{aligned} \tag{E6}$$

The deterministic contribution to the SDEs can be determined from the transitions via

$$x_1\left[F^{(0)}(\boldsymbol{x},\boldsymbol{e})\,-\,\varepsilon F^{(1)}(\boldsymbol{x},\boldsymbol{e})\right]= \sum_{\mu=1}^{u}\nu_{1\mu}T_\mu\left[R^2(\boldsymbol{x},\boldsymbol{e})^T+\boldsymbol{\nu}_\mu|(\boldsymbol{x},\boldsymbol{e})^T\right] \quad \text{(E7)}$$

$$x_2F^{(0)}(\boldsymbol{x},\boldsymbol{e})=\sum_{\mu=1}^{u}\nu_{2\mu}T_\mu\left[R^2(\boldsymbol{x},\boldsymbol{e})^T+\boldsymbol{\nu}_\mu|(\boldsymbol{x},\boldsymbol{e})^T\right] \quad \text{(E8)}$$

$$h_i(\boldsymbol{x},\boldsymbol{e})=\sum_{\mu=1}^{u}\nu_{(i+2)\mu}T_\mu\left[R^2(\boldsymbol{x},\boldsymbol{e})^T+\boldsymbol{\nu}_\mu|(\boldsymbol{x},\boldsymbol{e})^T\right]\,.$$

Notice that the relationship between Eqs. (E7) and (E8) is controlled by assumption 2. The correlations in the noise meanwhile are given by

$$\langle\eta_1(t)\eta_1(t')\rangle = \delta(t-t')\lim_{\varepsilon\to0}\sum_{\mu=1}^{u}\nu_{1\mu}^2T_\mu\left[R(\boldsymbol{x},\boldsymbol{e})+\boldsymbol{\nu}_\mu|(\boldsymbol{x},\boldsymbol{e})\right]\,, \quad \text{(E9)}$$

$$\langle\eta_2(t)\eta_2(t')\rangle = \delta(t-t')\lim_{\varepsilon\to0}\sum_{\mu=1}^{u}\nu_{2\mu}^2T_\mu\left[R(\boldsymbol{x},\boldsymbol{e})+\boldsymbol{\nu}_\mu|(\boldsymbol{x},\boldsymbol{e})\right]\,, \quad \text{(E10)}$$

$$\langle\eta_1(t)\eta_2(t')\rangle = 0\,, \quad \text{(E11)}$$

$$\langle\eta_i(t)\beta_j(t')\rangle = \delta(t-t')\lim_{\varepsilon\to0}\sum_{\mu=1}^{u}\nu_{i\mu}\nu_{(j+2)\mu}T_\mu\left[R(\boldsymbol{x},\boldsymbol{e})+\boldsymbol{\nu}_\mu|(\boldsymbol{x},\boldsymbol{e})\right] \quad \text{(E12)}$$

$$\langle\beta_i(t)\beta_j(t')\rangle = \delta(t-t')\lim_{\varepsilon\to0}\sum_{\mu=1}^{u}\nu_{(i+2)\mu}\nu_{(j+2)\mu}T_\mu\left[R(\boldsymbol{x},\boldsymbol{e})+\boldsymbol{\nu}_\mu|(\boldsymbol{x},\boldsymbol{e})\right]\,, \quad \text{(E13)}$$

at leading order in ε. The lack of noise correlation between the phenotypes, Eq. (E11), is a consequence of assumption 3. Assumption 2 allows us to rewrite Eqs. (E9) and (E10) as

$$\begin{aligned}\langle\eta_1(t)\eta_1(t')\rangle &= \delta(t-t')x_1H(\boldsymbol{x},\boldsymbol{e})\,,\\ \langle\eta_2(t)\eta_2(t')\rangle &= \delta(t-t')x_2H(\boldsymbol{x},\boldsymbol{e})\,.\end{aligned} \quad \text{(E14)}$$

An example of a system where this condition is not enforced is explored in Section F.

To begin our analysis of the SDEs, a quasi-neutral limit is considered in which $\varepsilon = 0$. Then the deterministic ODEs for the system (the SDEs in the limit $R\to\infty$) lead to a manifold of fixed points associated with the focus phenotypes. We now make two additional assumptions;

5. There exits a single stable, well behaved, manifold

6. This manifold is one-dimensional, and so can be paramaterized by a single variable

We then choose to parameterize the manifold in terms of x_1, which for clarity we label z on the CM. The CM is then defined by the set of equations

$$x_1=z\,,\qquad x_2=Z_2(z)\,,\qquad e_i=Z_{ei}(z)\quad\forall\, i=2,\dots N\,. \quad \text{(E15)}$$

The system dynamics are now entirely analogous to that of the public good model in the quasi-neutral limit. Deterministically, the system comes to rest at a point on the CM (which depends on the system's initial conditions) at which it stays indefinitely, and when stochasticity is included the system moves along the CM until one of the phenotypes fixates. A timescale separation is present so long as the composition of the population changes on a slower timescale to that of the collapse to the CM. In practice, the timescale of the collapse to the CM can be inferred from the eigenvalues of Eq. (E6) linearised about the CM. The magnitude of the smallest non-zero eigenvalue is indicative of the slowest component of collapse to the CM [34]. This should be much larger than the timescale at which the system moves along the CM, which is of order R^{-1} [35].

In order to implement the timescale separation, a non-linear projection is applied to the system which maps fluctuations back to the CM. This can be seen to be equivalent to transforming into the deterministically invariant variable whose existence is guaranteed by the existence of the CM [31], setting the dynamics in all other variables equal to zero, and evaluating the variables themselves on the CM. What form does this mapping take, in the quasi-neutral limit, for Eq. (E6)? Since the dynamical equations for the phenotypes take on the form of degenerate replicator equations in the limit $\varepsilon\to0$, the ratio x_1/x_2 is deterministically invariant, regardless of the other parameters. Therefore the non-linear mapping may be obtained by solving the following equation for z;

$$\frac{z}{Z_2(z)}=\frac{x_1}{x_2}\,,\quad\to\quad z=Y(x_1,x_2)\,. \quad \text{(E16)}$$

The resulting effective description for the quasi-neutral system on the CM can be denoted

$$\dot z=\frac{1}{R}\mathcal{S}(z)+\frac{1}{\sqrt{R}}\zeta(t)\,. \quad \text{(E17)}$$

Note that while the deterministic system evaluated on the CM had no drift dynamics, the reduced system may. Mathematically, this is a consequence of the fact that the equations are defined strictly in the Itō sense (from the underlying IBM) and therefore the normal rules of calculus do not apply. Instead, any nonlinear transformation induces a drift, in general given by

$$\mathcal{S}(z)=\frac{1}{2}\left[\sum_{ij}^{2}\left(\frac{\partial z}{\partial x_i\partial x_j}B_{ij}\right)+\sum_{ij}^{N}\left(\frac{\partial z}{\partial e_i\partial e_j}B_{eij}\right)\right]\Bigg|_{x_1=z,x_2=Z_2(z),e_i=Z_{ei}(z)} \quad \text{(E18)}$$

However, since the mapping z is independent of the ecosystem variables $\boldsymbol{e}$ (see Eq. (E16)), Eq. (E18) can be simplified to

$$\mathcal{S}(z) = \frac{1}{2}\sum_{ij}^{2}\left(\frac{\partial z}{\partial x_i \partial x_j}B_{ij}\right)\Bigg|_{x_1=z, x_2=Z_2(z), e_i=Z_{ei}(z)} . \tag{E19}$$

The form of the correlations in $\zeta(t)$ are now given by

$$\mathcal{B}(z) = \sum_{ij}^{2}\left(\left[\frac{\partial z}{\partial x_i}\right]_i\left[\frac{\partial z}{\partial x_j}\right]_j B_{ij}(\boldsymbol{x})\right)\Bigg|_{x=z, x_2=Z_2(z), e_i=Z_{ei}(z)} \tag{E20}$$

where once again we have taken advantage of the property $(dz/de_i) = 0$ for all i.

In this very general scenario, what inferences can we make about $\mathcal{S}(z)$? To answer this, it is convenient to return to our original SDEs, Eq. (E6), and implement the timescale separation in a different fashion. We begin by transforming into variables measuring the total size of the x_1 and x_2 population and the fraction of type x_1;

$$\begin{aligned} & N_T = x_1 + x_2\,, \quad f_1 = \frac{x_1}{x_1 + x_2}\,, \\ \rightarrow \quad & x_1 = f_1 N_T\,, \quad x_2 = N_T(1 - f_1)\,. \end{aligned} \tag{E21}$$

Applying this transformation, taking care to account for the impact of Itō calculus, we arrive at the following SDEs for the system;

$$\begin{aligned} \frac{\mathrm{d}f_1}{\mathrm{d}t} &= \frac{1}{2R^2}\sum_{i,j=1}^{2}\frac{\partial^2 f_1}{\partial x_i \partial x_j}B_{ij} + \frac{1}{R}\tilde{\eta}_1(t)\,, \\ \frac{\mathrm{d}N_T}{\mathrm{d}t} &= N_T F^{(0)}(\boldsymbol{x}, \boldsymbol{e}) + \frac{1}{R}\tilde{\eta}_2(t)\,, \\ \frac{\mathrm{d}e_i}{\mathrm{d}t} &= h_i(\boldsymbol{x}, \boldsymbol{e}) + \frac{1}{R}\tilde{\beta}_i(t)\,, \quad \forall\, i = 1, \ldots N\,. \end{aligned} \tag{E22}$$

By conducting the transformation, we immediately notice a few things. Most trivially, the forms of the noise correlations are now altered in all variables. Second, since the transformation into the variable N_T was linear, its governing SDE contains no noise-induced elements. Finally, the non-linear transformation into f_1 has resulted in a noise induced drift term. This drift term however is only dependent on the noise correlation structure between x_1 and x_2. Evaluating the dynamics for N_T and $\boldsymbol{e}$ on the CM and substituting in the remaining expressions from Eqs. (E11) and (E14), we obtain the following one-dimensional SDE for f_1;

$$\frac{\mathrm{d}f_1}{\mathrm{d}t} = \frac{1}{R}\tilde{\eta}_1(t)\,, \tag{E23}$$

where $\tilde{\eta}_1(t)$ is evaluated on the CM. There are no deterministic dynamics in our reduced dimension description of f_1. This is a consequence of assumptions 2 and 3. The equation for the fixation probability of phenotype X_1 given an initial *fraction* f_{10} on the CM, $Q(f_{10})$, is then, regardless of the noise form,

$$Q(f_{10}) = f_{10}\,. \tag{E24}$$

Crucially however, f_1 is evaluated on the CM, which may vary depending on the constitution of the population;

$$f_{10} = \frac{x_{10}}{x_{10} + Z_2(x_{10})}\,. \tag{E25}$$

If $[dZ_2(x_{10})/dx_{10}] < 1$, then the total phenotype population decreases with increasing x_{20}, and phenotype X_1 , has a larger invasion probability than X_2. From this we can infer that $\mathcal{S}(z)$ will be positive on average along the length of the CM;

$$\int_{z=0}^{N_1/R^2}\mathcal{S}(z)dz > 0\,. \tag{E26}$$

Therefore, the phenotype with the higher carrying capacity will be stochastically selected for in this quasi-neutral case, regardless of their interaction with the environment. We note once again that this result is in general dependent on assumption 2. If assumption 2 does not hold then there will be correlations between the fluctuations $\eta_1(t)$ and $\eta_2(t)$ and, rather than the equation for the time evolution of f_1 featuring no mean drift (as in Eq. (E23) there will be a noise induced drift term favoring one or other of the phenotypes. The exact form of this term will be highly dependent on the exact form of the interactions between the phenotypes, a full treatment of which lies outside the scope of this paper.

Now suppose that $\varepsilon > 0$, so that the system is non-neutral. Now there exists no CM. There is no line of deterministic fixed points, and therefore no invariant variable to project our variables on to and reduce the problem. However, under the assumption that ε is small there is still a separation of timescales. If ε is sufficiently small, the slow manifold (and the projection to it) can be approximated by the results from the quasi-neutral case (see Eqs. (E15) and (E16)), plus an ε correction. A perturbative analysis can thus be conducted, and, under the assumption the $\varepsilon \approx \mathcal{O}(R^{-2})$, at leading order we have

$$\dot{z} = -\varepsilon\mathcal{D}(z) + \frac{1}{R^2}\mathcal{S}(z) + \frac{1}{R}\bar{\eta}(t)\,. \tag{E27}$$

The form of $\mathcal{S}(z)$ is unchanged from Eq. (E18), while the new deterministic contribution to the drift takes the form

$$\mathcal{D}(z) = -\sum_{i=1}^{N}\left(\frac{dz}{dx_i}\frac{dx_i}{dt}\right)\Bigg|_{x_1=z, x_2=Z_2(z), \boldsymbol{e}=\boldsymbol{Z_e}(z)} . \tag{E28}$$

Once again however, the projection is simply a function of x_1 and x_2, and so

$$\begin{aligned} \mathcal{D}(z) = -\Bigg(& x_1 F^{(0)}(\boldsymbol{x})\frac{dz}{dx_1} + x_2 F^{(0)}(\boldsymbol{x})\frac{dz}{dx_2} \\ & - \varepsilon x_1 F^{(1)}\frac{dz}{dx_1}\Bigg)\Bigg|_{x_1=z, x_2=Z_2(z), \boldsymbol{e}=\boldsymbol{Z_e}(z)} . \end{aligned} \tag{E29}$$

Finally, we also know that in the limit $\varepsilon \to 0$ this deterministic contribution to the dynamics on the CM, $\mathcal{D}(z)$, should disappear. Therefore the first two terms in the above equation must cancel, leaving us with

$$\mathcal{D}(z) = \varepsilon z \left(F^{(1)}(\boldsymbol{x}) \frac{dz}{dx_1} \right) \Big|_{x_1 = z, x_i = Z_i(z)} . \tag{E30}$$

We now have a much simpler system to deal with. Say that $F^{(1)}(\boldsymbol{x})$ is strictly positive. Then this will be a term which consistently decreases the value of x_1. Based on physical arguments, we would expect that, regardless of the form of ζ, $\mathcal{D}(z)$ must be positive. We still require the exact form of z (see Eq. (E16)) to make analytic progress and specific predictions. Generally however, we have shown that $\mathcal{S}(z)$ will be positive so long as species X_1 has a larger carrying capacity (subject to the above conditions). A consideration of Eq. (E27) shows that even when the system is non-neutral, for sufficiently weak selection/small R, there will be a tradeoff between stochastic 'strength in numbers' and deterministic costs for high-density behavior.

Appendix F: Illustrating generality with reference to a complimentary systems: The stochastic Lotka-Volterra system

In Section B it was noted that deterministically the public good model reduces to a competitive Lotka-Volterra model under the elimination of the fast public good dynamics. However, it is important to note that though they may be deterministically equivalent at long times, due to alterations in the demographic noise structure the two systems have distinct behaviors. Despite this, the qualitative picture remains the same; for the quasi-neutral system, the fixation probability of each type is simply proportional to its initial fraction in the population, while when selection is introduced, there is playoff between stochastic and deterministic effects. To illustrate this, we investigate the stochastic Lotka-Volterra competition model (SLVC), derived from first principles.

In this section we analyze a stochastic Lotka-Volterra competition model using the methods developed in Section E. We assume a population composed of two phenotypes, X_1 and X_2, whose numbers in the system are measured by $\boldsymbol{n} = (n_1, n_2)$. The phenotypes are born, die and compete with each other. In particular, we define the system to be governed by the probability transition rates

$$\begin{aligned} T_1(n_1+1, n_2|n_1, n_2) &= b_1 n_1 , \\ T_2(n_1-1, n_2|n_1, n_2) &= d_1 n_1 + \tfrac{c_1}{R^2} n_1^2 + \tfrac{c_2}{R^2} n_1 n_2 , \\ T_3(n_1, n_2+1|n_1, n_2) &= b_2 n_2 , \\ T_4(n_1, n_2-1|n_1, n_2) &= d_2 n_2 + \tfrac{c_1}{R^2} n_1 n_2 + \tfrac{c_2}{R^2} n_2^2 . \end{aligned}$$

Together with Eq. (A2), this fully specifies the stochastic dynamics. Taking the limit of large R, we can once again obtain a mesoscopic description of the system;

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= x_1 \left((b_1 - d_1) - c_1 x_1 - c_2 x_2 \right) + \frac{1}{R} \eta_1(t), \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= x_2 \left((b_2 - d_2) - c_1 x_1 - c_2 x_2 \right) + \frac{1}{R} \eta_2(t), \end{aligned} \tag{F1}$$

where $\eta_i(t)$ have correlation structure Eq. (A8) with $B_{ij}(\boldsymbol{x})$ term given by

$$\begin{aligned} B_{11}(\boldsymbol{x}) &= x_1 \left((b_1 + d_1) + c_1 x_1 + c_2 x_2 \right) , \\ B_{22}(\boldsymbol{x}) &= x_2 \left((b_2 + d_2) + c_1 x_1 + c_2 x_2 \right) , \\ B_{12}(\boldsymbol{x}) &\equiv B_{21}(\boldsymbol{x}) = 0 . \end{aligned} \tag{F2}$$

Note that the noise structure is *not* the same as that in Eq. (E14); two phenotypes with an equal effective reproduction rate $b_1 - d_1 = b_2 - d_2$ have the same deterministic fitness, but distinct multiplicative noise. Phenotypes which are reproducing and dying more quickly are subject to greater noise as they have a larger rate of population turnover. We will however proceed to consider this more general scenario in order to illustrate what can happen when this assumption is not enforced. Finally, we impose a separation of timescales by setting

$$b_1 - d_1 = \tilde{b}(1 - \varepsilon) , \quad b_2 - d_2 = \tilde{b} . \tag{F3}$$

A CM thus exists if $\varepsilon = 0$, and an SM while ε is small. The parameter $\tilde{b}$ is an effective birth rate encompassing birth and death, while ε is a fitness cost paid by phenotype X_1 either in terms of a decreased birth rate, or increased death rate, relative to phenotype X_2.

In the case $\varepsilon = 0$, the system is quasi-neutral, and so a CM exists. The equation for the CM $x_2 = Z_2(x_1)$ (see Eq. (E15)) and its intersection with the boundaries $x_2 = 0$ and $x_1 = 0$, $K_1^{(0)}$ and $K_2^{(0)}$ respectively, are

$$\begin{aligned} Z_2(x_1) &= \frac{1}{c_2} \left(\tilde{b} - c_1 x_1 \right) , \\ K_1^{(0)} &= \frac{\tilde{b}}{c_1} , \quad K_2^{(0)} = \frac{\tilde{b}}{c_2} . \end{aligned} \tag{F4}$$

The parameters $K_1^{(0)}$ and $K_2^{(0)}$ give the frequency of X_1 and X_2 phenotypes in isolation. We assume that $c_2 > c_1$ and thus that phenotype X_1 exists at higher densities than phenotype X_2. Finally, the mapping from any point (x_1, x_2) to a coordinate $z = x_1$ on the CM is determined from Eq. (E16);

$$z = \frac{\tilde{b} x_1}{c_1 x_1 + c_2 x_2} . \tag{F5}$$

We can now obtain expressions for $\mathcal{D}(z)$, $\mathcal{S}(z)$ and $\mathcal{B}(z)$ directly from Eqs. (E30), (E18) and (E20);

$$\mathcal{D}(z) = -z \left(\tilde{b} - c_1 z \right) , \tag{F6}$$

$$\mathcal{S}(z) = \frac{2}{\tilde{b}^2} z \left(\tilde{b} - c_1 z \right) \left(c_2(\tilde{b} + d_2) - c_1(\tilde{b} + d_1) \right) , \tag{F7}$$

$$\begin{aligned} \mathcal{B}(z) = \frac{2}{\tilde{b}^2} z \left(\tilde{b} - c_1 z \right) \Big[z \Big(& c_2(\tilde{b} + d_2) \\ & - c_1(\tilde{b} + d_1) \Big) + \tilde{b}(d_1 + \beta) \Big] . \end{aligned} \tag{F8}$$

The equation can now be solved to calculate the fixation probability of phenotype X_1 along the CM. In terms of the initial fraction of X_1, f_1, we find

$$Q(f_1) = \frac{1-\chi(f_1)}{1-\chi(1)}\,,$$
$$\chi(f_1) = \left[\left(\frac{K_1^{(0)}}{d_1+\tilde{b}}\right)\left(\frac{d_1(1-f_1)+d_2f_1+\tilde{b}}{K_1^{(0)}(1-f_1)+f_1K_2^{(0)}}\right)\right]^{-\theta} \quad \text{(F9)}$$

where θ is a parameter given by

$$\theta = \left(1+\frac{K_1^{(0)}K_2^{(0)}R^2\tilde{b}\varepsilon}{K_2^{(0)}(d_1+\tilde{b})-K_1^{(0)}(d_2+\tilde{b})}\right). \quad \text{(F10)}$$

Let us consider the special case $\varepsilon = 0$. The fixation probability then becomes

$$Q(f_1)|_{\varepsilon=0} = \frac{f_1(d_2+\tilde{b})}{d_1(1-f)+d_2f+\tilde{b}}\,. \quad \text{(F11)}$$

The species with the lower death rate (and death rate, since $\tilde{b}$ is fixed), has a greater probability of fixation than the species with the higher birth rate/death rate. This insight, made in [9, 10], is a result of the higher levels of noise experienced by the phenotype with the high birth and death rates. This makes it easier for the longer lived phenotype, (lower birth/death rates), to invade and fixate. For the purposes of this paper, we ignore such effects in order to focus on systems in which the carrying capacity of the phenotypes alone is responsible for the differences in noise experienced by the phenotypes on the CM/SM.

To this end, we now focus on the case $b_1 = b_2 \equiv b$, $d_1 = d_2 \equiv d$. In this case, $Q(f_1)|_{\varepsilon=0} = f_1$, and $Q(f_1)$ in general becomes

$$Q(f_1) = \frac{1-\chi(f_1)}{1-\chi(1)}\,,$$
$$\chi(f_1) = \left(\frac{K_1^{(0)}}{K_1(1-f_1)+f_1K_2^{(0)}}\right)^{-\theta}, \quad \text{(F12)}$$

where θ is now given by

$$\theta = \left(1+\frac{K_1^{(0)}K_2^{(0)}R^2(b-d)\varepsilon}{(K_2^{(0)}-K_1^{(0)})b}\right). \quad \text{(F13)}$$

The invasion probabilities ϕ_1 and ϕ_2 meanwhile are given by

$$\phi_1 = Q(N_2^{-1})\,, \quad \phi_2 = 1-Q(1-N_1^{-1})\,. \quad \text{(F14)}$$

We can use the above expressions to obtain an approximate value for the maximum cost to birth rate that can be paid in order that selection reversal is observed. Assuming N_1^{-1} and N_2^{-1} are of order ε and Taylor expanding in ε, we find the cost to birth must obey

$$\frac{1}{N_2}\left(\frac{b}{b-d}\right)\left(1-\frac{N_2}{N_1}\right) > \varepsilon \quad \text{(F15)}$$

for the direction of selection to be reversed. This is analogous to Eq. (8) in the main text.

Appendix G: Order of magnitude parameter estimates

In this section we seek an illustrative set of parameters for use in the model in order emphasis that the insights developed are biologically reasonable. We wish to obtain order of magnitude estimates for the set of parameters, b, p_x, p_u, r, δ, κ, R, m and D. We choose the yeast *Saccharomyces cerevisiae* as our model organism. While our model is more physically realistic than many mathematical public good models, we note that there are still choices that must be made in relating this physical system to our general framework.

Our model is constructed such that the uptake of one constituent of the public good, Q, by a phenotype, results in a reproduction event. In the context of *S. cerevisiae*, the type Q is thus shorthand for the amount of invertase that must be present in the system to break down sucrose into sufficient glucose for a reproduction event of the yeast. Let us define σ to be the scaling between n_q and the total number of invertase molecules, such that the number of invertase molecules is σn_q. In order to understand the relationship between our model parameters and physically measurable parameters, we begin by considering a simplified ODE system of our model.

$$\frac{dx}{dt} = x(b+rq-\kappa x)\,, \quad \text{(G1)}$$
$$\frac{dq}{dt} = px-\delta q\,. \quad \text{(G2)}$$

While the total number of discrete invertase constituents is $n_q \approx R^2q$, the total number of invertase molecules is $R^2\sigma q$. Let θ be a measure of the number of invertase molecules, such that $\theta = \sigma q$. The ODEs in this more natural variable read

$$\frac{dx}{dt} = x(b+\frac{r}{\sigma}\theta-\kappa x)\,, \quad \text{(G3)}$$
$$\frac{d\theta}{dt} = \sigma px-\delta\theta\,. \quad \text{(G4)}$$

The decay rate δ is independent of the number of molecules which make up an invertase constituent Q, so we can take experimental measurements of the invertase molecular decay rate as values for δ. Meanwhile the molecular invertase production rate and reproduction rate due to invertase take on scaled forms of the parameters in our original ODEs;

$$r_{mol} = \frac{r}{\sigma}\,, \quad \text{(G5)}$$
$$p_{mol} = \sigma p\,. \quad \text{(G6)}$$

While measurements of p_{mol} are obtainable in the literature (see Table S.1), our estimation of r_{mol} is complicated by the fact that it is an effective parameter. It must capture the increase in the reproductive rate due to invertase, which in reality is coupled to both the reaction rate of invertase and sucrose into glucose, as well as the

uptake rate of glucose by yeast and the energy conversion to reproduction. We do however know the typical range of yeast reproduction rates. Let us define λ_{exp} as the yeast reproduction rate as measured experimentally. In turn, let λ_{eff} be the effective per capita reproduction rate of yeast in the model;

$$\lambda_{\text{eff}} = b + rq\,. \tag{G7}$$

The yeast reproduction rate clearly depends on the amount of public good in the system, typically varying from

$$\begin{aligned} &\lambda_{\text{eff}} = b \qquad (\text{all non}-\text{producers}) \\ \text{to} \qquad &\lambda_{\text{eff}} = b\frac{\delta\kappa}{\delta\kappa - p_x r} \qquad (\text{all producers})\,. \end{aligned} \tag{G8}$$

In reality, the reproduction rate of yeast in a system without any invertase is effectively zero; we have assumed some baseline birth rate for convenience in the model, which could be physically interpreted as being associated with an exogenous glucose concentration in the system. We assume that this is typically low, such that b is small, while the yeast approaches its maximum reproductive rate when it consists entirely of producers.

The parameter κ controls death due to crowding. For simplicity this is the only form of death in the model. This choice leads, perhaps unnaturally, to the non-producers (who exist at typically lower densities) having a much smaller death rate than producers. For the parameters chosen however, we obtain per- capita death rates on the order of an hour for producers, and ten hours for non-producers. The parameter R meanwhile measures the assumed spatial interaction scale. It determines the typical number of individuals on each patch. We can use this value to infer the size of a patch. Denoting the diameter of a yeast cell as L_c, and assuming that the hyper-producing cells in the stationary state can be packed on a grid, the size of each patch, L_p, can be approximated by

$$L_p = L_c\sqrt{N_u}\,, \tag{G9}$$

$$= L_c R\sqrt{K_u}\,. \tag{G10}$$

The parameters m and D are effective migration and diffusion rates in our model. To map these physical parameters these must be in turn scaled by the patch length. The public good diffusion rate must also be scaled by σ, which maps the discrete amount of invertase constituents Q to the number of invertase molecules. Denoting m_{exp} and D_{exp} the physical migration and public good diffusion rate rates, it can be shown that [28]

$$D_{exp} = \sigma L_p^2 D\,, \qquad m_{exp} = L_p^2 m\,. \tag{G11}$$

The parameter choices which follow from these calculations are summarized in Table S.2.

When considering the parameter choices summarized in Table S.2, it is important to make a final point. While both the approximations we have employed, the system size expansion in Section A and the fast-variable elimination in Section B, rely formally on R^2 being large and ε being small ($\mathcal{O}(R^{-2}) > \varepsilon$), in practical terms the procedures are relatively robust to this restriction. Indeed, throughout the body of the main text, $R = 4$, while ε is varied on the interval $[0, 0.1]$. In fact we find that the approximate analytic expression we obtain for the invasion probabilities of the phenotypes, Eq. (C4), describes the results obtained from simulation well, up to $\varepsilon = R^{-2}$, as illustrated in Fig. 8.

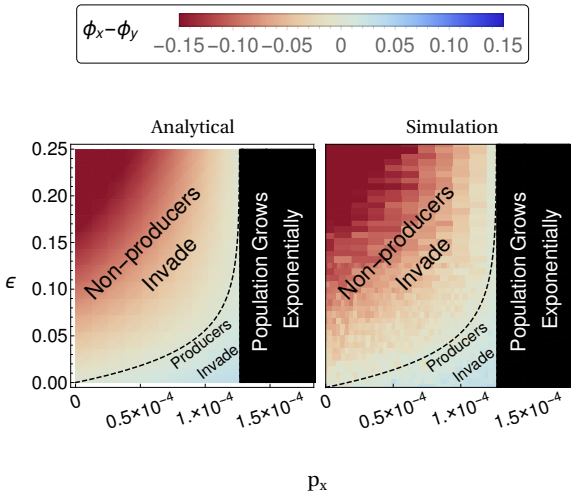


FIG. 8. Figure illustrating the larger range of values for the parameter ε over which the approximation Eq. (C4) is accurate. Parameters are given in Table S.2, with the exception of p_x and ε which are varied. Note that this figure is similar to Fig. 2, plotted in the main text, but plotted over a grater range of ε and $\phi_x - \phi_y$. The parameter region plotted in black is that for which $\kappa\delta > rp_x$.

In terms of the system size expansion, this robustness can in part be explained by the fact that the typical population sizes (N_x, N_y and N_q) are proportional to R. For populations of fixed size N, SDEs for the system can be obtained by means of a Taylor expansion of the master equation (for example, Eq. (A2)) as a series in $1/N$. A crucial feature of the system we are concerned with here however, is that population sizes may vary, and so this technique is unavailable. Instead we conduct an expansion in the interaction scale R, which is proportional to the mean population size. Though R may not be a large number itself, increasing R leads to an associated increase in population size (see Table S.2). In turn, this leads to terms of higher order in the Taylor expansion of the master equation becoming subdominant [47], justifying the truncation which leads to Eq. (A4). In contrast, the resilience of the fast-variable elimination approximation to such large values of ε is surprising.

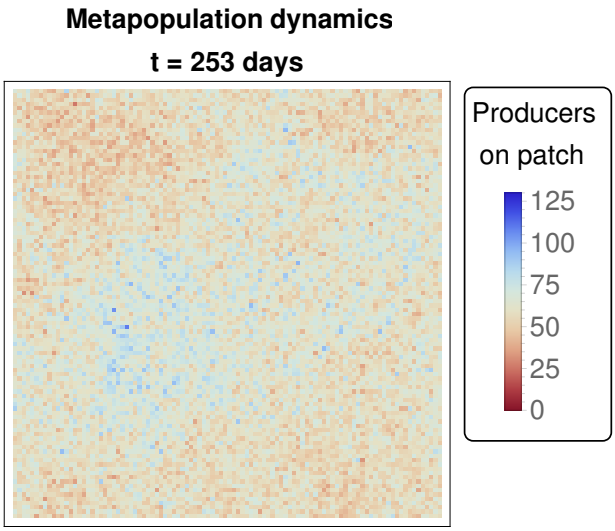
Experimental parameters		
Experimental Parameter	Value	Description
p_{mol}	$0.46\,\mathrm{mol}\,s^{-1}$	Production rate of a molecule of invertase per producing yeast cell [26].
δ	$2\times10^{-3}\,\mathrm{mol}\,s^{-1}$	Estimated efficacy decay rate of invertase (see [49], Fig 5.)
λ_{exp}	$0.31-0.5\,hr^{-1}$	Yeast reproduction rate in producing population [50, 51]
ε_{exp}	0.06	Cost of public good production to yeast reproduction rate [50]
D_{exp}	$100\,\mu m^2 s^{-1}$	Diffusion rate of invertase molecules estimated in [41]
L_c	$3\,\mu m$	Cell length physical approximation [41].

TABLE S.1. List of experimental parameters obtained from literature.

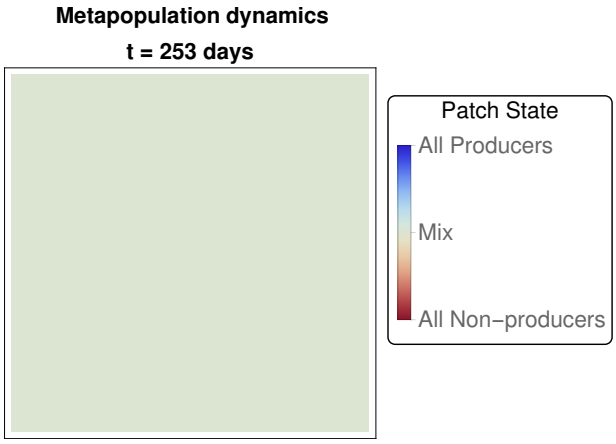
Illustrative parameter choices with justifications		
Parameter	Value	Justification
σ	4000	Assumed parameter. Presence of 4000 invertase molecules required for yeast reproduction.
p_y	0	True non-producer does not produce invertase.
p_x	$1.14\times10^{-4}\,s^{-1}$, $(0.41\,hr^{-1})$	Experimental value of molecular invertase production rate (see Table S.1) scaled by σ (see Eq. (G6)).
p_u	$1.2\times10^{-4}\,s^{-1}$ $(0.43\,hr^{-1})$	Leads to factor 1.7 increase in the steady state invertase from producing to hyper-producing population, consistent with [33].
b	$6.94\times10^{-6}\,s^{-1}$ $(0.025\,hr^{-1})$	Small baseline yeast birthrate assumed.
r	$1.58\times10^{-5}\,s^{-1}$ $(0.057\,hr^{-1})$	Chosen so as to give per-capita yeast reproduction rate $(b+rq)\approx\lambda_{exp}$ when system entirely producers (see Table S.1).
δ	$0.002\,s^{-1}$	Taken from experimentally measured values. (see Table S.1)
κ	$1\times10^{-6}\,s^{-1}$	Suggested parameter for illustrating effects in paper; restricted by $\delta\kappa > p_i r$, $i=x,y,u$.
R	2	Suggested parameter for illustrating effects in paper.
ε	0.06	Taken from experiments (see Table S.1)
N_y	28	See Eq. (B11).
N_x	302	See Eq. (B11).
N_u	499	See Eq. (B11) for N_x and substitute p_u for p_x.
L_p	$67\,\mu m$	See Eq. (G10).
m	$3.4\times10^{-7}\,s^{-1}$	Yields a migration to birth-rate ratio between $m/b=4.9\times10^{-2}$ (all non-producers) and $m/(b+rq)=4.5\times10^{-3}$ (all producers).
D	$2.22\times10^{-5}\,s^{-1}$	Obtained using experimental value D_{exp} from Table S.1 and Eq. (G11).

TABLE S.2. List of parameters used in the simulation, with the exception of p_x, p_u, ε, m and D which are varied.

Appendix H: Movies



Movie 1. Still from Movie S1: Movie of a simulation of the metapopulation public good model on a 100×100 grid ($C = 10^4$). Colors indicate the number of producers on each patch; patches with a small number of producers are colored red while patches with a large number of producers are colored blue. Parameters used are $p_x = 1 \times 10^{-4}$, $\varepsilon = 0.02$, $m = 3.7 \times 10^{-5}$ and the remaining parameters taken from Table S.2. With these parameters, $N_y \approx 28$ and $N_x \approx 129$. Initial conditions are a single producer and non-producer on each patch. Large numbers of producers on a patch are correlated with low numbers of non-producers on the same patch. The space-averaged dynamics of this simulation are given in the main text, Fig. 4. Counter to the deterministic prediction, the number of producers increases with time, while the number of non-producers decreases.



Movie 2. Still from Movie S2: Movie showing the distribution of homogeneous non-producing patches (red), homogeneous producing patches (blue) and heterogeneous mixed patches (gray-green) in the simulation of the metapopulation public good model given in Video S1. For the majority of the observation time, every patch contains a heterogeneous mix of producers and non-producers. Homogeneous producer patches only begin to emerge as producers approach fixation in the system.

[1] Hofbauer J, Sigmund K (1998) *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge).

[2] Fisher RA (1930) *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford).

[3] Wright S (1931) Evolution in Mendelian populations. *Genetics* 16:97–159.

[4] Nowak MA (2006) *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, Cambridge, Massachusetts).

[5] McKane AJ, Newman TJ (2005) Predator-prey cycles from resonant amplification of demographic stochasticity. *Phys. Rev. Lett.* 94:218102.

[6] Butler T, Goldenfeld N (2009) Robust ecological pattern formation induced by demographic noise. *Phys. Rev. E* 80:030902.

[7] Hallatschek O, Hersen P, Ramanathan S, Nelson DR (2007) Genetic drift at expanding frontiers promotes gene segregation. *PNAS* 104:19926–19930.

[8] Rossberg AG, Rogers T, McKane AJ (2013) Are there species smaller than 1 mm? *Proc. Roy. Soc. B* 280:20131248.

[9] Parsons T, Quince C (2007) Fixation in haploid populations exhibiting density dependence I: The non-neutral case. *Theor. Pop. Biol.* 72:121–135.

[10] Lin YT, Kim H, Doering CR (2012) Features of fast living: On the weak selection for longevity in degenerate birth-death processes. *J. Stat. Phys.* 148:646–662.

[11] Gunawardena J (2014) Time-scale separation - Michaelis and Mentens old idea, still bearing fruit. *FEBS Journal* 281:473–488.

[12] Parsons T, Quince C (2007) Fixation in haploid populations exhibiting density dependence II: The quasi-neutral case. *Theor. Pop. Biol.* 72:468–479.

[13] Chotibut T, Nelson DR (2015) Evolutionary dynamics with fluctuating population sizes and strong mutualism. *Phys. Rev. E* 92:022718.

[14] Constable GWA, McKane AJ (2015) Models of genetic drift as limiting forms of the Lotka-Volterra competition model. *Phys. Rev. Lett.* 114:038101.

[15] Parsons TL, Quince C, Plotkin JB (2010) Some consequences of demographic stochasticity in population genetics. *Genetics* 185:1345–1354.

[16] Kogan O, Khasin M, Meerson B, Schneider D, Myers CR (2014) Two-strain competition in quasi-neutral stochastic disease dynamics. *Phys. Rev. E* 90:042149.

[17] Houchmandzadeh B, Vallade M (2012) Selection for altruism through random drift in variable size populations. *BMC Evol. Biol.* 12:61.

[18] Houchmandzadeh B (2014) Fluctuation driven fixation of cooperative behavior. *Biosystems* 127:60–66.

[19] Reznick D, Bryant MJ, Bashey F (2002) r- and k-selection revisited: The role of population regulation in life-history evolution. *Ecology* 83:1509–1520.

[20] Nowak MA, Sasaki A, Taylor C, Fudenberg D (2004) Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428:646–650.

[21] Rice SH (2004) *Evolutionary Theory* (Sinauer Associates, Sunderland, Massachusetts).

[22] Crow JF, Kimura M (1970) *An Introduction to Population Genetics Theory* (The Blackburn Press, New Jersey).

[23] Gore J, Youk H, van Oudenaarden A (2009) Snowdrift game dynamics and facultative cheating in yeast. *Nature* 459:253–256.

[24] Hauert C, Holmes M, Doebeli M (2006) Evolutionary games and population dynamics: maintenance and cooperation in public goods games. *Proc. R. Soc. B* 273:2565–2570.

[25] Huang W, Hauert C, Traulsen A (2015) Stochastic game dynamics under demographic fluctuations. *Proc. Natl. Acad. Sci.* 112:9064–9069.

[26] Koschwanez JH, Foster KR, Murray AW (2011) Sucrose utilization in budding yeast as a model for the origin of undifferentiated multicellularity. *PLoS Biol.* 9:e1001122.

[27] Kümmerli R, Brown SP (2010) Molecular and regulatory properties of a public good shape the evolution of cooperation. *Proc. Natl. Acad. Sci.* 107:18921–18926.

[28] Gardiner CW (2009) *Handbook of Stochastic Methods* (Springer, Berlin).

[29] Black AJ, McKane AJ (2012) Stochastic formulation of ecological models and their applications. *Trends Ecol. Evol.* 27:337–345.

[30] Parsons TL, Rogers T (2015) Dimension reduction via timescale separation in stochastic dynamical systems. *arXiv* p 1510.07031.

[31] Arnold L (2003) *Random Dynamical Systems*, Springer Monographs in Mathematics (Springer, Berlin).

[32] Maclean RC, Fuentes-Hernandez A, Greig D, Hurst LD, Gudelj I (2010) A mixture of "cheats" and "cooperators" can enable maximal group benefit. *PLoS Biol.* 21:e1000486.

[33] Maclean RC, Brandon C (2008) Stable public goods cooperation and dynamic social interactions in yeast. *J. Evol. Biol.* 21:1836–1843.

[34] Constable GWA, McKane AJ, Rogers T (2013) Stochastic dynamics on slow manifolds. *J. Phys. A: Math. Theor.* 46:295002.

[35] Constable GWA, McKane AJ (2014) Fast-mode elimination in stochastic metapopulation models. *Phys. Rev. E* 89:032141.

[36] Behar H, Brenner N, Ariel G, Louzoun Y (2016) Fluctuations-induced coexistence in public good dynamics.

[37] Behar H, Brenner N, Louzoun Y (2014) Coexistance of productive and non-productive populations by fluctuation-driven spatio-temporal patterns. *Theor. Popul. Biol.* 96:20–29.

[38] Nowak MA, Tarnita CE, Antal T (2009) Evolutionary dynamics in structured populations. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365:19–30.

[39] Wakano JY, Hauert C (2011) Pattern formation and chaos in spatial ecological public goods games. *J. Theor. Biol.* 268:30–38.

[40] Nowak MA, May RM (1992) Evolutionary games and spatial chaos. *Nature* 359:826–829.

[41] Allen B, Gore J, Nowak MA (2013) Spatial dilemmas of diffusible public goods. *eLife* 2:01169.

[42] Waxman D, Gavrilets S (2005) 20 questions on adaptive dynamics. *Journal of Evolutionary Biology* 18:1139–1154.

[43] Pianka ER (1970) On r- and k-selection. *The American Naturalist* 104:592–597.

[44] Lin YT, Kim H, Doering CR (2015) Demographic stochasticity and evolution of dispersion I. spatially homogeneous environments. *J. Math. Biol.* 70:647.

[45] Lin YT, Kim H, Doering CR (2015) Demographic stochasticity and evolution of dispersion II. spatially inhomogeneous environments. *J. Math. Biol.* 70:679.

[46] Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* 22:403–434.

[47] van Kampen NG (2007) *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam).

[48] Risken H (1989) *The Fokker-Planck Equation* (Springer, Berlin).

[49] Gomez L, Ramirez HL, Cabrera G, Simpson BK, Villalonga R (2008) Immobilization of invertasechitosan conjugate on hyaluronic-acid-modified chitin. *J. Food Biochem.* 32:264 – 277.

[50] Sanchez A, Gore J (2013) Feedback between population and evolutionary dynamics determines the fate of social microbial populations. *PLoS Biol.* 11:e1001547.

[51] Snoep JL, Mrwebi M, Schuurmans JM, M. RJ, de Mattos M. J. T (2009) Control of specific growth rate in Saccharomyces cerevisiae. *Microbiology* 155:1699 – 1707.