

PNAS



Supplementary Information for

Reproductive variance can drive behavioral dynamics

Guocheng Wang, Qi Su, Long Wang, and Joshua B. Plotkin

Qi Su, Long Wang, Joshua B. Plotkin.

E-mail: qisu1991@sas.upenn.edu, longwang@pku.edu.cn, jplotkin@sas.upenn.edu

This PDF file includes:

Supplementary text

Figs. S1 to S9

References for SI reference citations

Supporting Information Text

Contents

1	Basic assumptions and population model	2
2	Methods	4
A	Fixation probability	4
B	Intuition for the effects of demographic stochasticity	6
3	Examples: some explicit birth-death processes	7
4	Extensions	9
A	General games	9
A.1	ODE-based analysis	10
A.2	Stationary distribution	11
A.3	Games with multiple strategies	12
B	Dynamics in slow-growing populations	12
B.1	Fixation probability	13
B.2	Extinction and persistence	13
	Appendix	15
	Supplementary figures	17

1. Basic assumptions and population model

We use a continuous-time, discrete-state Markov process to describe births and deaths in a population of replicating individuals. We study a family of population models by making the following axiomatic assumptions about the birth process:

Assumption 1.1 *Let the stochastic process $X_t^{(i)}$ denote the number of offspring that individual i produces in time $0 \rightarrow t$. We assume that the stochastic process satisfies the following four properties:*

- (1) $X_0^{(i)} = 0$;
- (2) *Independent increment: $X_{t+\tau}^{(i)} - X_t^{(i)}$ is independent of $X_t^{(i)}$;*
- (3) *Stationary increment: $X_{t+\tau}^{(i)} - X_t^{(i)}$ and $X_\tau^{(i)}$ have the same distribution;*
- (4) *Non-decreasing: $X_{t+\tau}^{(i)} - X_t^{(i)} \geq 0$.*

Given these properties, the mean and variance of $X_t^{(i)}$ are both linear functions of t , and the offspring distribution (the number of offspring within unit of time) is homogeneous and independent of time. We use ξ_i to denote the offspring distribution. To describe the distribution of ξ_i we make the following axiomatic assumptions about its expectation and variance:

Assumption 1.2 *The mean and variance of ξ_i are given by*

- (1) $\mathbb{E}(\xi_i) = B + s\pi_i$;
- (2) $\text{Var}(\xi_i) = \delta_1 B + \delta_2 s\pi_i$.

Here, the parameter B denotes the baseline of birth rate, π_i denotes individual i 's payoff (obtained from pairwise social interactions in the population), and s denotes the selection intensity. A larger value of s means that payoffs have a stronger effect on individual's birth process. In this paper, we adopt the common assumption that the selection intensity is weak ($s \ll 1$).

The expected number of offspring in one unit of time, $\mathbb{E}(\xi_i)$, is called the fitness of individual i . In the regime of weak selection we can expand $\mathbb{E}(\xi_i)$ as a Taylor series and truncate it to first order in π_i . There may be a number of different birth models for which this truncation yields part (1) of Assumption 1.2. Note that in a classical discrete-time model of reproduction, such as the Moran process, fitness is often assumed to be an exponential function of payoff, i.e., $\exp(B + s\pi_i)$. This is consistent with our definition of fitness $\mathbb{E}(\xi_i) = B + s\pi_i$ in continuous time.

The space of models defined by the axioms above includes models in which the birth process is not Poisson, but is rather over-dispersed. There are many ways to produce over-dispersion, where the mean and variance are correlated in different ways (1, 2), such as quasi-Poisson model (variance is proportional to mean), mixed-effects Poisson model and negative binomial model (variance is a quadratic function of mean). Here, we stipulate only a general form of correlation between variance and mean: $\text{Var}(\xi_i) = f(\mathbb{E}(\xi_i)) = f(B + s\pi_i)$. In the regime of weak selection $s \ll 1$, we can expand the variance to first order in s , which yields (2) in Assumption 1.2. The parameter δ_1 measures how the variance component in the offspring distribution scales with the baseline birth rate, which is the same for all individuals. Whereas the parameter δ_2 measures how the variance component of the offspring distribution scales with the current payoff of individual i , which may vary across individuals. The case $\delta_1 = \delta_2 = 1$ simplifies to a classic Poisson birth process. Whereas if $\delta_1 > 1$, the birth processes is over-dispersed (because we work in the regime $s \ll 1$).

Suppose there are N individuals of a given phenotype. Thus these N individuals have the same but independent birth rates. The overall birth rate of this phenotype is then

$$\sum_{i=1}^N \xi_i, \quad [1]$$

which converges to a normal distribution when N is large, by the central limit theorem. We approximate the Markov process by a diffusion equation, which depends only on the first two moments of ξ_i and is expected to be a good approximation in large enough populations. In addition, we perform explicit simulations of processes, and compare their results to the predictions derived by the diffusion approximation, in Section 3.

We assume death events follow a Poisson process governed by two terms, for each phenotype. One term arises from a constant baseline death rate, denoted by D . The other term corresponds to deaths caused by competition among individuals for limited resources, which introduces a carrying capacity on the entire population. Specifically, the competition process is modelled by a reaction process



which takes place with constant rate λ and X and Y are any two individuals. Thus, the death rate caused by competition is λ times the population size.

Given these assumptions, the population is described by a continuous-time discrete state Markov process, which can be approximated by a diffusion equation. We start by using the traditional donation game as an example, to produce payoff expressions and derive the corresponding diffusion equation. Consider a population composed of x cooperators and y defectors. Each individual plays games with all other players (including self) and obtains an average payoff. Each cooperator provides a benefit b to his opponent and pays a cost c , but a defector contributes nothing and pays no cost. The payoff matrix for a pairwise interaction is as follows:

$$\begin{array}{cc} & \begin{array}{cc} \text{C} & \text{D} \end{array} \\ \begin{array}{c} \text{C} \\ \text{D} \end{array} & \begin{pmatrix} b-c & -c \\ b & 0 \end{pmatrix}. \end{array} \quad [3]$$

Summing over all pairwise interactions, the total payoff for a cooperator and defector are

$$\pi_C = \frac{x}{x+y}b - c, \quad [4a]$$

$$\pi_D = \frac{x}{x+y}b. \quad [4b]$$

We assume that in a short time interval Δt , the number of births of cooperators is Δx^+ , and the number of deaths of cooperators is Δx^- . Thus the increment of cooperators is $\Delta x = \Delta x^+ - \Delta x^-$. The same applies to defectors, y . Then, we obtain the following expressions for mean change in state of short time Δt

$$\mathbb{E}[\Delta x(t)|x(t), y(t)] = x(t)[B + s\pi_C - D - \lambda(x(t) + y(t))]\Delta t, \quad [5a]$$

$$\mathbb{E}[\Delta y(t)|x(t), y(t)] = y(t)[B + s\pi_D - D - \lambda(x(t) + y(t))]\Delta t. \quad [5b]$$

Since deaths follow a Poisson process, the variance of Δx^- equals to its mean, namely $Dx\Delta t + \lambda x(x+y)\Delta t$. Moreover, Δx^+ is independent of Δx^- , which leads to the following expressions for the variance in state change of short time Δt

$$\begin{aligned} \mathbb{E}[\Delta x(t)^2|x(t), y(t)] &\approx \text{Var}[\Delta x(t)^+ - \Delta x(t)^-] = \text{Var}[\Delta x(t)^+] + \text{Var}[\Delta x(t)^-] \\ &= x(t)(\delta_1 B + \delta_2 s\pi_C)\Delta t + Dx(t)\Delta t + \lambda x(t)(x(t) + y(t))\Delta t + o(\Delta t), \end{aligned} \quad [6a]$$

$$\begin{aligned} \mathbb{E}[\Delta y(t)^2|x(t), y(t)] &\approx \text{Var}[\Delta y(t)^+ - \Delta y(t)^-] = \text{Var}[\Delta y(t)^+] + \text{Var}[\Delta y(t)^-] \\ &= y(t)(\delta_1 B + \delta_2 s\pi_D)\Delta t + Dy(t)\Delta t + \lambda y(t)(x(t) + y(t))\Delta t + o(\Delta t). \end{aligned} \quad [6b]$$

Thus, we obtain

$$\lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}[\Delta x(t)|x(t) = x, y(t) = y]}{\Delta t} = x[B + s\pi_C - D - \lambda(x + y)], \quad [7a]$$

$$\lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}[\Delta y(t)|x(t) = x, y(t) = y]}{\Delta t} = y[B + s\pi_D - D - \lambda(x + y)], \quad [7b]$$

$$\lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}[\Delta x(t)^2|x(t) = x, y(t) = y]}{\Delta t} = (\delta_1 B + \delta_2 s\pi_C)x + Dx + \lambda x(x + y), \quad [7c]$$

$$\lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}[\Delta y(t)^2|x(t) = x, y(t) = y]}{\Delta t} = (\delta_1 B + \delta_2 s\pi_D)y + Dy + \lambda y(x + y), \quad [7d]$$

$$\lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}[\Delta x(t)\Delta y(t)|x(t) = x, y(t) = y]}{\Delta t} = 0. \quad [7e]$$

In what follows, for simplicity we use α to denote $B - D$, which represents the baseline net growth rate of the population, independent of payoff effects. Then we can then use the following diffusion equation to approximate the Markov process(3):

$$dx = x[\alpha + s\pi_C - \lambda(x + y)]dt + \sqrt{(\delta_1 B + \delta_2 s\pi_C)x + Dx + \lambda x(x + y)}dW_t^{(1)}, \quad [8a]$$

$$dy = y[\alpha + s\pi_D - \lambda(x + y)]dt + \sqrt{(\delta_1 B + \delta_2 s\pi_D)y + Dy + \lambda y(x + y)}dW_t^{(2)}, \quad [8b]$$

where $W_t^{(1)}$ and $W_t^{(2)}$ are two independent Wiener processes, and the diffusion terms are interpreted in the sense of Itô.

2. Methods

A. Fixation probability. To investigate the dynamics of population size and cooperator frequency, we introduce a coordinates transformation, setting $p = x/(x + y)$ and $n = x + y$. Applying Itô's lemma

$$dp = \frac{\partial p}{\partial x}dx + \frac{\partial p}{\partial y}dy + \frac{1}{2}\frac{\partial^2 p}{\partial x^2}(dx)^2 + \frac{1}{2}\frac{\partial^2 p}{\partial y^2}(dy)^2, \quad [9a]$$

$$dn = \frac{\partial n}{\partial x}dx + \frac{\partial n}{\partial y}dy + \frac{1}{2}\frac{\partial^2 n}{\partial x^2}(dx)^2 + \frac{1}{2}\frac{\partial^2 n}{\partial y^2}(dy)^2, \quad [9b]$$

to Eq. 8 we obtain

$$\begin{aligned} dp = & -scp(1-p)dt + \frac{1-p}{n}\sqrt{x(\delta_1 B + \delta_2 s\pi_C + D + \lambda n)}dW_t^{(1)} \\ & - \frac{p}{n}\sqrt{y(\delta_1 B + \delta_2 s\pi_D + D + \lambda n)}dW_t^{(2)} - \frac{xy(\delta_1 B + \delta_2 s\pi_C + D + \lambda n)}{n^3}\left(dW_t^{(1)}\right)^2 \\ & + \frac{xy(\delta_1 B + \delta_2 s\pi_D + D + \lambda n)}{n^3}\left(dW_t^{(2)}\right)^2, \end{aligned} \quad [10a]$$

$$\begin{aligned} dn = & [n\alpha + s(b-c)pn - \lambda n^2]dt + \sqrt{x(\delta_1 B + \delta_2 s\pi_C + D + \lambda n)}dW_t^{(1)} \\ & + \sqrt{y(\delta_1 B + \delta_2 s\pi_D + D + \lambda n)}dW_t^{(2)}. \end{aligned} \quad [10b]$$

Remembering $(dW_t)^2 = dt$ and omitting terms of order $O(s)$ in the diffusion term of Eq. 10a (more accurately, we further assume $\delta_1 B \gg \delta_2 s$ such that $\delta_2 s\pi_C$ and $\delta_2 s\pi_D$ can be omitted), we obtain

$$dp = scp(1-p)\left(\frac{\delta_2}{n} - 1\right)dt + \frac{1-p}{n}\sqrt{x(\delta_1 B + D + \lambda n)}dW_t^{(1)} - \frac{p}{n}\sqrt{y(\delta_1 B + D + \lambda n)}dW_t^{(2)}, \quad [11a]$$

$$dn = [n\alpha + s(b-c)pn - \lambda n^2]dt + \sqrt{x(\delta_1 B + D + \lambda n)}dW_t^{(1)} + \sqrt{y(\delta_1 B + D + \lambda n)}dW_t^{(2)}. \quad [11b]$$

We work in the regime of $s \ll \alpha$. Here the population growth rate is mainly determined by the baseline birth and death rates α . If $\alpha > 0$, the population will grow logistically until it reaches the carrying capacity. If $\alpha < 0$, the population will eventually perish. The case of $\alpha = 0$ is a critical case, where the growth rate is totally determined by payoffs. In this paper, we only discuss the case of $\alpha > 0$.

We first consider a corresponding deterministic system, which serves as a reference to this stochastic system. In the deterministic system we neglect stochasticity (variance) in the number of birth and death events per time interval. For the birth process, when $\delta_1 = \delta_2 = 0$, the variance becomes zero and thus the birth process simplifies to a deterministic process. The death process is modelled as a Poisson process. Each individual dies at rate $D + \lambda n$. And the overall death rate is $Dn + \lambda n^2$.

Thus, within one unit of time the expected number of individuals who die obeys a Poisson distribution with mean $Dn + \lambda n^2$ and variance $Dn + \lambda n^2$. In the corresponding deterministic system we ignore the variance of the death process, as well, which means we assume the number of individuals who die is not stochastic, but strictly equal to $Dn + \lambda n^2$. Given these assumptions, the stochastic term of Eq. 8 vanishes and the equation simplifies to an ordinary differential equation:

$$dx = x [\alpha + s\pi_C - \lambda(x + y)] dt \quad [12a]$$

$$dy = y [\alpha + s\pi_D - \lambda(x + y)] dt. \quad [12b]$$

By a parameter transformation we obtain

$$dp = -scp(1-p)dt \quad [13a]$$

$$dn = [n\alpha + s(b-c)pn - \lambda n^2]dt. \quad [13b]$$

Here, dp is always negative, which means a cooperator will never be favored by selection.

In the regime $s \ll \alpha$, the system features a timescale separation. n equilibrates much more quickly than p . Therefore, given an arbitrary initial configuration p_0 and n_0 , before p changes at all the dynamical system rapidly converges to the slow manifold defined by $dn = 0$, or more explicitly,

$$n = \frac{\alpha + s(b-c)p}{\lambda} \approx \frac{\alpha}{\lambda}. \quad [14]$$

The slow manifold is in fact the carrying capacity of the environment, denoted by M . It reflects the maximum number of individuals that the environment can sustain. For the stochastic system, the system still shows the features of fast-slow dynamics (Fig. S2)(4). The population size n will quickly converges to M , and then fluctuate around it.

Since the population reaches the carrying capacity rapidly, and the population size then fluctuate around M in subsequent evolution, we can use M to replace n in Eq. 11a. Remembering $D + \lambda M = B$, we obtain a one-dimensional diffusion system describing the evolution of cooperator frequency on the slow manifold:

$$dp = scp(1-p) \left(\frac{\delta_2}{M} - 1 \right) dt + \sqrt{\frac{(\delta_1 + 1)Bp(1-p)}{M}} \left(\sqrt{1-p}dW_t^{(1)} - \sqrt{p}dW_t^{(2)} \right). \quad [15]$$

The solution of Eq. 15 is a Markov process with the infinitesimal generator given by

$$\mathcal{L}f = scp(1-p) \left(\frac{\delta_2}{M} - 1 \right) \frac{\partial f}{\partial p} + \frac{(\delta_1 + 1)Bp(1-p)}{2M} \frac{\partial^2 f}{\partial p^2}. \quad [16]$$

For Eq. 8, $x = 0$ and $y = 0$ are two absorbing states. Given the initial number of cooperators x_0 and defectors y_0 , for some time t , if $x(t) > 0$ and $y(t) = 0$ (i.e. $p = 1$) is satisfied, we say cooperators have fixed (5). The probability that cooperators will fix is called fixation probability, denoted ρ . Given the initial frequency of cooperators p_0 ($p_0 = x_0/(x_0 + y_0)$), the fixation probability of cooperators $\rho(p_0)$ is the solution of

$$\begin{cases} \mathcal{L}\rho = 0, \\ \rho(1) = 1, \\ \rho(0) = 0. \end{cases} \quad [17]$$

The solution of this equation can be calculated explicitly (3), which is

$$\rho(p_0) = \frac{\int_0^{p_0} S(x)dx}{\int_0^1 S(x)dx}, \quad [18]$$

where

$$S(x) = \exp \left(\int -\frac{2E(p)}{D(p)} dp \right), \quad [19a]$$

$$E(p) = scp(1-p) \left(\frac{\delta_2}{M} - 1 \right), \quad [19b]$$

$$D(p) = \frac{(\delta_1 + 1)Bp(1-p)}{M}. \quad [19c]$$

Using this method, we obtain the fixation probability:

$$\rho(p_0) = \frac{\exp \left[\frac{2sc}{(\delta_1 + 1)B} (M - \delta_2)p_0 \right] - 1}{\exp \left[\frac{2sc}{(\delta_1 + 1)B} (M - \delta_2) \right] - 1} \approx p_0 - \frac{sc}{(\delta_1 + 1)B} (M - \delta_2)p_0(1-p_0). \quad [20]$$

If $s \rightarrow 0$, $\rho(p_0) \rightarrow p_0$, which means that fixation probability in the absence of selection is equal to the initial frequency of cooperators. Furthermore, if $2scM/(\delta_1 B + B) - 2\delta_2 sc/(\delta_1 B + B) < 0$, i.e.

$$M < \delta_2, \quad [21]$$

the fixation probability will exceed that of neutral drift, i.e. $\rho(p_0) > p_0$, which means that cooperation is favored by natural selection.

Our analysis in the main text and above is based on an approximation by separation of timescales. Here we briefly discuss the accuracy of this approximation. Suppose the system starts at state (n_0, p_0) , with $n_0 < M$. When the population size hits the slow manifold $n = M$ we denote the fraction of cooperators as p_s . The precise fixation probability is then

$$\rho(p_0, n_0) = p_s - \frac{sc}{(\delta_1 + 1)B} (M - \delta_2) p_s (1 - p_s). \quad [22]$$

For $\alpha \gg s$, n equilibrates much more quickly than p . And so in our analysis by separation of timescales, so we use the approximation that p_s equals p_0 . However, in the full SDE system with finite α , p_s will not equal p_0 exactly; there will be some slight deviation in p by the time $n = M$. Given Eq. 11, the drift term of dp is

$$scp(1 - p) \left(\frac{\delta_2}{n} - 1 \right) \quad [23]$$

Under analysis by timescale separation we derived $\delta_2 = M$ as a critical value that determines whether cooperation is favored or not. When we choose $\delta_2 = M$, we note that the drift term above becomes $scp(1 - p)(M/n - 1)$. Before the system reaches the slow manifold, then, the drift term is always positive. which means p is expected to increase slightly while approaching the slow manifold in the full SDE. Thus, p_s is expected to be slightly larger than p_0 . As a result (Eq. 22), without making the approximation of timescale separation the fixation probability is actually slightly larger than p_0 , so that our approximation Eq. 21 is an over-estimate. Likewise, the actual critical value of δ_2 to favor cooperation, without approximation by timescale separation, is slightly lower than M , which can be seen visually in Fig. 2 in the main text.

B. Intuition for the effects of demographic stochasticity. In Fig. 3 of the main text, we illustrate how fluctuations taking the trajectories off the slow manifold quickly return to the slow manifold, travelling along one of the the fast manifolds. The position where the trajectories return to the slow manifold is expected to have a deviation from the original starting point, which yields an effective advective force on the cooperator frequency p along the slow manifold. Here, we provide detailed calculation for this advective force.

As summarized in Fig. 3 of the main text, we suppose the initial number of cooperators and defectors are x and y , lying on the slow manifold $x + y = M$. As shown in Fig. 3, due to the noise the trajectory may fluctuate to point $x + \Delta x$ and $y + \Delta y$ within a short time Δt . Then, it will return along a fast manifold to the slow manifold at point (x', y') . As a result there is an effective advective force on the number of cooperators (i.e. $x' - x$) along the slow manifold. We now derive the expectation of $x' - x$.

As summarized in Fig. 3B in the main text, we can obtain an expression for the return point:

$$x' = (x + y) \frac{x + \Delta x}{x + y + \Delta x + \Delta y}. \quad [24]$$

Then, we have

$$\begin{aligned} \mathbb{E}[x' - x] &= \mathbb{E} \left[\frac{x + \Delta x}{1 + \frac{\Delta x + \Delta y}{x + y}} - x \right] \\ &\approx \mathbb{E} \left[(x + \Delta x) \left(1 - \frac{\Delta x + \Delta y}{x + y} + \frac{(\Delta x + \Delta y)^2}{(x + y)^2} \right) - x \right] \\ &= \mathbb{E} \left[-\frac{y}{(x + y)^2} \Delta x^2 + \frac{x}{(x + y)^2} \Delta y^2 \right]. \end{aligned} \quad [25a]$$

Here, $(\Delta x, \Delta y)$ obeys a two-dimensional Gaussian distribution. According to Eq. 7, we have

$$\mathbb{E}[\Delta x^2] \approx \text{Var}(\Delta x) = x(\delta_1 B + \delta_2 s \pi_C + D + \lambda(x + y)) \Delta t, \quad [26a]$$

$$\mathbb{E}[\Delta y^2] \approx \text{Var}(\Delta y) = y(\delta_1 B + \delta_2 s \pi_D + D + \lambda(x + y)) \Delta t. \quad [26b]$$

Substituting Eq. 26 into Eq. 25, we have

$$\mathbb{E}[x' - x] = \frac{xy}{(x + y)^2} \delta_2 s (\pi_D - \pi_C) \Delta t. \quad [27]$$

Thus, when $\pi_C < \pi_D$, the advective force induced by noise tends to increase the number of cooperators. When $\pi_D < \pi_C$, the number of cooperators is expected to decrease, and thus defectors are expected to increase.

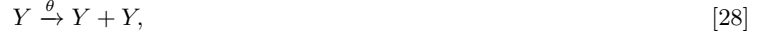
Furthermore, since $(\Delta x, \Delta y)$ obeys a two-dimensional Gaussian distribution, the shape of the fluctuation is an ellipse. The lengths of the two principle axes represent the standard deviation of Δx and Δy (square root of the variance). We can choose the four endpoints of the ellipse to illustrate the direction of advective force (Fig. S4).

3. Examples: some explicit birth-death processes

We have analyzed a large class of models by stipulating axiomatic forms for the mean and variance in the birth process, and studying it by approximation using a diffusion equation. Here, for concreteness sake, we specify some explicit discrete-state/continuous time processes that satisfy our axioms (Assumption 1.1 and Assumption 1.2), and we demonstrate that our analysis predicts their behavior.

Poisson process

Most prior studies depict birth events often through a reaction equation



where Y is an arbitrary individual and θ is the birth rate. This assumption produces a birth process that is Poisson with intensity θ . Within a time interval $[t, t + \Delta t]$, the offspring number of the individual obeys a Poisson distribution with expectation $\theta_i \Delta t$ and variance $\theta_i \Delta t$. As usual, the variance is equal to the mean in this classic case.

In our more general family of models we consider the number of offspring ξ_i produced by individual i within one unit of time, which is assumed to satisfy

$$\mathbb{E}(\xi_i) = B + s\pi_i = \theta_i, \quad [29a]$$

$$\text{Var}(\xi_i) = \delta_1 B + \delta_2 s\pi_i = \theta_i. \quad [29b]$$

Our family of models thus contains the classical Poisson birth process in the case case $\delta_1 = \delta_2 = 1$. In this case, the condition that cooperation is favored becomes to

$$M < 1, \quad [30]$$

which can never be achieved since the carrying capacity can never be lower than 1. And so, under our family of models, a strictly Poisson birth process can never favor cooperation.

Compound Poisson process

In Poisson process there is only one offspring produced in each reproduction event. However, for many species in nature, a large number of offspring are produced simultaneously in each reproduction event. We call the number of offspring produced (instantaneously) in a single reproduction event the "litter size". In general the litter size is stochastic. In this section we describe how to model such a birth process using a compound Poisson process.

We use M_t to denote the times of reproduction events from 0 to t and we assume that M_t is a Poisson process with intensity θ . And in each reproduction event, the litter size Z obeys a distribution with mean μ and variance σ^2 . Therefore, the total number of offspring in $[0, t]$ is

$$X(t) = \sum_{k=1}^{M(t)} Z_k. \quad [31]$$

We can derive the mean and variance of $X(t)$.

$$\begin{aligned} \mathbb{E}X(t) &= \mathbb{E}[\mathbb{E}[X(t)|M(t)]] \\ &= \mathbb{E}[\mathbb{E}[M(t)\mu]] \\ &= \theta\mu t, \end{aligned} \quad [32a]$$

$$\begin{aligned} \text{Var}(X(t)) &= \mathbb{E}[X(t)^2] - (\mathbb{E}[X(t)])^2 \\ &= \mathbb{E}[\mathbb{E}[X(t)^2|M(t)]] - \theta^2\mu^2 t^2 \\ &= \mathbb{E}[M(t)\sigma^2 + M(t)^2\mu^2] - \theta^2\mu^2 t^2 \\ &= \theta(\sigma^2 + \mu^2)t. \end{aligned} \quad [32b]$$

Thus, $X(t)$ is also a stochastic process with stationary and independent increments that satisfies Assumption 1.1. From Assumption 1.2, we obtain

$$\mathbb{E}(\xi_i) = B + s\pi_i = \theta_i\mu_i, \quad [33a]$$

$$\text{Var}(\xi_i) = \delta_1 B + \delta_2 s\pi_i = \theta_i(\mu_i^2 + \sigma_i^2). \quad [33b]$$

Here, the litter size Z can be any random variable over non-negative integers, with an arbitrary probability distribution. In this paper we consider two specific cases: Z obeys a Poisson distribution, or a negative binomial distribution.

Case 1: A Poisson litter size

First we study the case where Z has the Poisson distribution. For a player with payoff π_i , we assume that M_t is a Poisson process with parameter θ_i , and Z obeys a Poisson distribution with parameter μ_i . Then, Eq. 33 gives that

$$B + s\pi_i = \theta_i\mu_i, \quad [34a]$$

$$\delta_1 B + \delta_2 s\pi_i = \theta_i(\mu_i^2 + \mu_i). \quad [34b]$$

The reproduction process of an individual is totally controlled by the two parameters, θ_i and μ_i . On the other hand, the reproduction process is determined by $s\pi_i$. Thus, θ_i and μ_i are functions of $s\pi_i$. In the regime of weak selection ($s \ll 1$), any function of $s\pi_i$ can be expanded in a Taylor series around $s = 0$. Neglecting higher order terms in s leads to θ_i and μ_i linear in π_i . Thus, we can assume

$$\theta_i = \theta_0 + k_\theta s\pi_i, \quad [35a]$$

$$\mu_i = \mu_0 + k_\mu s\pi_i. \quad [35b]$$

Substituting Eq. 35 into Eq. 34, we can obtain

$$B + s\pi_i = \theta_0\mu_0 + (\theta_0 k_\mu + \mu_0 k_\theta) s\pi_i + o(s), \quad [36a]$$

$$\delta_1 B + \delta_2 s\pi_i = \theta_0(\mu_0^2 + \mu_0) + [k_\theta(\mu_0^2 + \mu_0) + \theta_0(2\mu_0 k_\mu + k_\mu)] s\pi_i + o(s). \quad [36b]$$

Comparing the both sides of Eq. 36 yields

$$B = \theta_0\mu_0, \quad [37a]$$

$$1 = \theta_0 k_\mu + \mu_0 k_\theta, \quad [37b]$$

$$\delta_1 B = \theta_0(\mu_0^2 + \mu_0), \quad [37c]$$

$$\delta_2 = k_\theta(\mu_0^2 + \mu_0) + \theta_0(2\mu_0 k_\mu + k_\mu). \quad [37d]$$

There are four equations and four unknowns ($\theta_0, \mu_0, k_\theta, k_\mu$), and so this system is solvable. We can solve that:

$$\theta_0 = \frac{B}{\delta_1 - 1}, \quad [38a]$$

$$\mu_0 = \delta_1 - 1, \quad [38b]$$

$$k_\theta = \frac{(2\delta_1 - \delta_2) - 1}{(\delta_1 - 1)^2}, \quad [38c]$$

$$k_\mu = \frac{\delta_2 - \delta_1}{B}. \quad [38d]$$

It is worth noting that this construction is meaningful only when θ_i and μ_i are positive. Because $s \ll 1$, we expect that θ_i and μ_i are positive only when θ_0 and μ_0 are positive, which requires $\delta_1 > 1$. This means that for any parameters δ_1, δ_2 , and B , provided $\delta_1 > 1$, we can construct such a Poisson-Poisson process that satisfies Assumption 1.2. Note also that a birth process based on Poisson-Poisson process always naturally leads to over-dispersion in the number of offspring.

Case 2: A negative binomial litter size

Now we study the case where Z follows a negative binomial Poisson distribution. For a player with payoff π_i , we assume that M_t is a Poisson process with parameter θ_i , and Z obeys a negative binomial distribution with parameters q_i and m ($q_i \in (0, 1)$ and $m \in \mathbb{N}^*$). The negative binomial distribution has mean $q_i m / (1 - q_i)$ and variance $q_i m / (1 - q_i)^2$. Since m is an integer and cannot change continuously with π_i , we assume that all individuals share an identical m .

Then, Assumption 1.2 and Eq. 32 give

$$B + s\pi_i = \theta_i \left(\frac{q_i m}{1 - q_i} \right), \quad [39a]$$

$$\delta_1 B + \delta_2 s\pi_i = \theta_i \left(\frac{q_i^2 m^2}{(1 - q_i)^2} + \frac{q_i m}{(1 - q_i)^2} \right). \quad [39b]$$

Similarly, we further assume that θ_i and q_i are linear in π_i , that is

$$\theta_i = \theta_0 + k_\theta s\pi_i, \quad [40a]$$

$$q_i = q_0 + k_q s\pi_i. \quad [40b]$$

Substituting Eq. 40 into Eq. 39 and using similar techniques, we can solve to find

$$\theta_0 = \frac{B(1 + m)}{(\delta_1 - 1)m}, \quad [41a]$$

$$q_0 = \frac{\delta_1 - 1}{\delta_1 + m}, \quad [41b]$$

$$k_\theta = \frac{(2\delta_1 - \delta_2 - 1)(m + 1)}{(\delta_1 - 1)^2 m}, \quad [41c]$$

$$k_q = \frac{(m + 1)(\delta_2 - \delta_1)}{B(\delta_1 + m)^2}. \quad [41d]$$

Similarly, this construction is meaningful only when θ_0 is positive and q_0 is in $[0, 1]$, which implies $\delta_1 > 1$. And so we see that birth process based on Poisson-negative binomial process always leads to over-dispersion. Given these equations (Eq. 38 and 41), we can always choose parameters of the compound Poisson process that satisfy our general stipulations on the mean and variance in the total offspring produced per unit time (Assumption 2), provided $\delta_1 > 1$ and $\delta_2 > 0$.

We can compare Monte-Carlo simulations of these explicit population processes (discrete state, continuous time) to the analytical prediction for the fixation probability that we derived from a stochastic differential equation (Eq. 20). We find good agreement between the individual-based simulations and analytic approximations, for carrying capacities as small as $M = 100$ or $M = 200$ (Fig. S3). Note that in both cases shown in Fig. S3, for sufficiently large δ_2 we have $k_\theta < 0$ and $k_m > 0$ or $k_q > 0$. In other words, higher payoffs reduce the rate of birth events but increase the mean litter size per birth event; and when these effects are strong enough, then selection favors cooperation.

Here in our construction, we have assumed m is fixed and the same for all individuals. If m can vary for different individuals, the construction will be more complicated and not unique. Moreover, we have only considered two examples of compound Poisson process here, but many other construction can be derived similarly. And so Assumption 1.2 is very general axiom so that the space of models we analyze covers a wide range of birth processes.

4. Extensions

A. General games. So far, we have analyzed the classic donation games, to explore the effects of demographic stochasticity on the evolution of cooperation. To explore more general results, we extend our model to general two-player/two-action games. A general two-player two-action game has the following payoff structure

$$\begin{array}{c} \text{C} \quad \text{D} \\ \text{C} \quad \left(\begin{array}{cc} a & b \\ c & d \end{array} \right) \\ \text{D} \end{array}.$$

Here, the two strategies are still called cooperation (C) and defection (D), which are now just generic terms that carry no implied meaning. When two cooperators interact, they both receive payoff a . When a cooperator encounters a defector, the cooperator receives b while the defector receives c . Mutual defection brings d to both players.

If there are x players employing cooperation and y players employing defection, their payoffs are respectively

$$\pi_C = \frac{xa + yb}{x + y}, \quad [42]$$

$$\pi_D = \frac{xc + yd}{x + y}. \quad [43]$$

As before, we make the parameter transformation $p = x/(x + y)$ and $n = x + y$. Then, under the assumption of $\alpha \gg s$, we can still separate the timescale of n and p , which gives

$$dp = sp(1-p) \left[\left(1 - \frac{\delta_2}{M}\right) (b-d + (a-b-c+d)p) \right] dt + \sqrt{\frac{(\delta_1+1)Bp(1-p)}{M}} \left(\sqrt{1-p}dW_t^{(1)} - \sqrt{p}dW_t^{(2)} \right), \quad [44]$$

where M is the carrying capacity given by $M = \alpha/\lambda$.

A.1. ODE-based analysis. If the carrying capacity is large, the diffusion term is very small, and thus has little influence on the dynamics. Thus, to provide an rough intuitive interpretation, we first analyze Eq. 44 by omitting its diffusion term altogether, which simplifies to an ordinary differential equation (ODE). That is

$$dp = sp(1-p) \left[\left(1 - \frac{\delta_2}{M}\right) (b-d + (a-b-c+d)p) \right] dt. \quad [45]$$

If we ignore s and let $\delta_2 = 0$ in Eq. 45, it actually degenerates to the classic replicator equation.

We denote Eq. 45 as $dp = f(p)dt$ in what follows. This equation has three equilibrium points:

$$p = 0, \quad [46a]$$

$$p = 1, \quad [46b]$$

$$p = p^* = \frac{d-b}{a-b-c+d}. \quad [46c]$$

These equilibrium points are meaningful only if they are located in the domain $[0, 1]$. Games can be classified into three scenarios in terms of the number of equilibrium points within $[0, 1]$ and their stabilities. We will analyze the effects of demographic noise for the three scenarios respectively.

Dominance games:

For a dominance game, a player's best choice of strategy does not depend on the opponent's strategy. Without loss of generality, we suppose $a < c$ and $b < d$ such that defection is always the best choice (e.g., prisoner's dilemma).

Since the definition domain of p is $[0, 1]$, in this case, $p^* \notin [0, 1]$. Thus, Eq. 45 has only two fixed points: $p = 0$ and $p = 1$.

If $\delta_2 < M$, we have

$$\left. \frac{df(p)}{dp} \right|_{p=0, \delta_2 < M} < 0 \quad \text{and} \quad \left. \frac{df(p)}{dp} \right|_{p=1, \delta_2 < M} > 0. \quad [47]$$

Thus, only $p = 0$ (i.e. defectors take over the population) is stable. However, we find that

$$\left. \frac{df(p)}{dp} \right|_{\delta_2 < M} \cdot \left. \frac{df(p)}{dp} \right|_{\delta_2 > M} < 0, \quad [48]$$

which means $\delta_2 > M$ can alter the stability of equilibrium points. If $\delta_2 > M$, $p = 1$ becomes stable but $p = 0$ is unstable, which implies cooperators will take over the population.

Coexistence games:

If $c > a$ and $b > d$, this game is a coexistence game. The best choice for an individual is to choose the opposite strategy of her opponent's. In this case, all of the three points (Eq. 46) are in $[0, 1]$. Hence the system has three equilibrium points.

When $\delta_2 < M$, we have

$$\left. \frac{df(p)}{dp} \right|_{p=0, \delta_2 < M} > 0, \quad \left. \frac{df(p)}{dp} \right|_{p=1, \delta_2 < M} > 0, \quad \text{and} \quad \left. \frac{df(p)}{dp} \right|_{p=p^*, \delta_2 < M} < 0 \quad [49]$$

Thus, only $p = p^*$ is stable, which means cooperators and defectors will coexist in the equilibrium state.

But for $\delta_2 > M$, $p = 0$ and $p = 1$ becomes stable but $p = (d-b)/(a-b-c+d)$ is unstable, which means cooperators or defectors will finally take over the population. Cooperators and defectors can never coexist.

Coordination games:

If $a > c$ and $d > b$, the game is a coordination game. The best choice is to choose the same strategy as your opponent. Similar to the coexistence game, all of the three equilibrium points exist. However, given $\delta_2 < M$, the interior point (p^*) is no longer stable and the two points on the boundary ($p = 0$ and $p = 1$) are stable. Thus, the system will eventually converge to full-cooperator or full-defector state according to the initial configuration.

If $\delta_2 > M$, the coordination game can be transformed to a coexistence game. Specifically, p^* becomes stable and another two points are unstable, which means cooperators and defectors will coexist.

Generally, under the condition of $\delta_2 > M$, the changes on dynamics can be viewed as doing a transformation on the payoff matrix. For any games, it can transform the payoff matrix to its opposite one, i.e.,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow \begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}.$$

Thus, compared to the classical outcome, which occurs for $\delta_2 < M$, we have the following effects when $\delta_2 > M$: for a dominance game the dominant strategy is swapped; coexistence games are transformed to coordination games; a coordination games are transformed to coexistence games.

A.2. Stationary distribution. The analysis above is obtained by omitting the stochastic term, reducing the analysis to an ODE. If we retain the stochastic term, Eq. 45 is a Markov process that has two absorbing states $p = 0$ and $p = 1$. Thus, this Markov process is not ergodic and the stationary distribution is not unique. A technical approach to studying this is to assume the boundaries are reflecting, so that no states are absorbing (3), conditioned on $n > 0$. Mechanistically, this means that when the number of cooperators (defectors) becomes zero, a new cooperator (defector) arises instantly in the population. Thus the definition domain of p becomes $[1/M, 1 - 1/M]$ approximately. Similar techniques are also used in (6). This approach makes the Markov process ergodic and the stationary distribution (denoted by $v^*(p)$) is unique. A frequency p with greater density in stationary distribution means that trajectories spend more time in the neighborhood of p . The stationary distribution $v^*(p)$ is the solution of the following Fokker-Planck equation:

$$-\frac{d}{dp}\mathcal{A}(p)v^*(p) + \frac{1}{2}\frac{d^2}{dp^2}\mathcal{B}(p)v^*(p) = 0 \quad [50]$$

where $\mathcal{A}(p)$ and $\mathcal{B}(p)$ is given by

$$\mathcal{A}(p) = sp(1-p)\left(1 - \frac{\delta_2}{M}\right)[b-d + (a-b-c+d)p] \quad [51a]$$

$$\mathcal{B}(p) = \frac{(\delta_1 + 1)Bp(1-p)}{2M} \quad [51b]$$

The solution of this equation can be expressed explicitly (3):

$$v^*(p) = \frac{\mathcal{N}}{\mathcal{B}(p)} \exp\left[2\int_0^p \frac{\mathcal{A}(p')}{\mathcal{B}(p')} dp'\right], \quad [52]$$

where \mathcal{N} is a normalization constant such that

$$\int_{1/M}^{1-1/M} v^*(p) dp = 1. \quad [53]$$

By some basic manipulations, we obtain the stationary distribution. That is

$$v^*(p) = \frac{\mathcal{N}M}{(\delta_1 + 1)Bp(1-p)} \exp\left[\frac{2sM}{(\delta_1 + 1)B}\left(1 - \frac{\delta_2}{M}\right)\left[(b-d)p + \frac{1}{2}(a-b-c+d)p^2\right]\right] \quad [54]$$

In the following analyse we omit \mathcal{N} since it does not affect the shape of the distribution. We find that Eq. 54 is the product of two components:

$$v_1(p) = \frac{M}{(\delta_1 + 1)Bp(1-p)} \quad [55a]$$

$$v_2(p) = \exp\left[\frac{2sM}{(\delta_1 + 1)B}\left(1 - \frac{\delta_2}{M}\right)\left[(b-d)p + \frac{1}{2}(a-b-c+d)p^2\right]\right]. \quad [55b]$$

Here, $v_1(p)$ is an U-shape distribution, where the boundaries $p = 1$ and $p = 0$ have the largest probability density. For $v_2(p)$, since the exponential function is monotonic increasing, the shape of $v_2(p)$ is determined by the quadratic function inside. If $\delta_2 < M$, for the prisoner's dilemma (dominance game), $v_2(p)$ achieves its extremum at $p = 0$. For a coordination game or a coexistence game, $v_2(p)$ achieves its extremum in the interior ($p = p^*$) or on the boundary ($p = 0$ and $p = 1$) respectively. If $\delta_2 > M$, the maximum of $v_2(p)$ in the case of $\delta_2 < M$ is now turned to be minimum, and vice versa.

if $(\delta_1 + 1)B$ is small, the shape of $v^*(p)$ is mainly determined by $v_2(p)$ because exponential function is dominant. Thus, we can see the humps of the stationary distribution are in accordance with the stable equilibria in the deterministic analysis. And large δ_2 can alter the position of the humps in stationary distribution, which confirms that large δ_2 can transform the dynamics of a game. However, if $(\delta_1 + 1)B$ is large, $v^*(p)$ is dominated by $v_1(p)$. The stationary distribution is always U-shape, so all games have similar properties with coordination games. This kind of transformation is not found in ODE-based analysis.

Thus, we can conclude that for a coexistence game, the previously condition we derived for transformation into a coordination game, $\delta_2 > M$, is actually conservative, since δ_1 also works to transform the game into a coordination game. However, for a coordination game, the condition $\delta_2 > M$ is liberal, since the heterogeneity in variance (large δ_2) should offset the effects of δ_1 first when it tries to transform the game into a coexistence game.

A.3. Games with multiple strategies. All prior analyses have focused on games with only two possible actions. Here we extend our analysis to the case of games with more than two strategies, which includes the famous rock-paper-scissors game with non-transitive payoff structure. For a two-player game with m strategies, suppose its payoff matrix is

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{pmatrix}.$$

Let n_i denote the number of individuals who adopt strategy i (the population size is $n = \sum n_i$), and p_i denote the frequency of strategy i . This gives $p_i = n_i/n$. Collecting all frequencies, we get a vector $\mathbf{p} = [p_1, \dots, p_n]$. The payoff of individuals with strategy i is

$$\pi_i = (\mathbf{A}\mathbf{p})_i, \quad [56]$$

and its growth equation is

$$dn_i = n_i[\alpha + s\pi_i - \lambda n]dt + \sqrt{n_i(\delta_1 B + \delta_2 s\pi_i + D + \lambda n)}dW_t^{(i)}. \quad [57]$$

Applying Itô's lemma

$$dp_i = \sum_{k=1}^m \frac{\partial p_i}{\partial n_k} dn_k + \frac{1}{2} \sum_{k=1}^m \frac{\partial^2 p_i}{\partial n_k^2} (dn_k)^2, \quad [58]$$

and separating the timescale of n and p (assuming $\alpha \gg s$), we obtain the stochastic differential equation for p_i :

$$dp_i = s \left(1 - \frac{\delta_2}{M}\right) p_i (\pi_i - \bar{\pi}) dt + \sqrt{\frac{(\delta_1 + 1)B p_i}{M}} dW_t^{(i)} - p_i \sum_{j=1}^m \sqrt{\frac{(\delta_1 + 1)B p_j}{M}} dW_t^{(j)} \quad [59]$$

where $\bar{\pi} = \sum p_i \pi_i$ is the average payoff. This is a $n - 1$ dimensional system. For large M , the diffusion term can also be ignored. This shows that for games with multiple strategies, when $\delta_2 > M$, the sign of the drift term can also be changed. And so once again, similar to the case of two-action games, we find that large heterogeneity in offspring variance due to payoff (large δ_2) can reverse the direction of evolution.

Using similar techniques, we can also analyze the stationary distribution with reflecting boundaries. Suppose the stationary distribution is $v^*(\mathbf{p}) = v^*(p_1, p_2, \dots, p_{n-1})$. The corresponding Fokker-Planck equation is

$$-\sum_{i=1}^{n-1} \frac{\partial}{\partial p_i} A_i(\mathbf{p}) v^*(\mathbf{p}) + \frac{1}{2} \sum_{i,j=1}^{n-1} \frac{\partial^2}{\partial p_i \partial p_j} B_{ij}(\mathbf{p}) v^*(\mathbf{p}) = 0, \quad [60]$$

where

$$A_i(\mathbf{p}) = \left(s - \frac{s\delta_2}{M} \right) p_i (\pi_i - \bar{\pi}) \quad [61]$$

$$B_{ij}(\mathbf{p}) = \begin{cases} -\frac{(\delta_1+1)B}{M} p_i p_j & i \neq j \\ \frac{(\delta_1+1)B}{M} p_i (1 - p_i) & i = j \end{cases}. \quad [62]$$

Unfortunately, it is difficult to obtain an explicit solution for this equation. And so we resort to simulations to show that large offspring variance can qualitatively change the directions of evolution and the stationary distribution (Fig. S5).

B. Dynamics in slow-growing populations. In all our analyses above we have assumed that the net baseline growth rate, $\alpha = B - D$, is sufficiently large compared to selection intensity s , so that the population rapidly approaches carrying capacity. In this setting the analysis can be simplified to a one-dimensional system (for two-strategy games), by separation of timescales. By contrast, in this section we focus on the case in which the net baseline growth rate of the population is small. In particular, we study the case when baseline growth has the same order as selection, $\alpha = O(s)$. In this regime the population grows very slowly and fixation of one type or another typically occurs long before reaching carrying capacity. The key question still remains: can large offspring variance qualitatively change the evolutionary outcome? We analyze this question in the case of the donation game.

B.1. Fixation probability. Since the system can no longer be simplified to a single dimension along the slow manifold, the fixation probability will now depend on the initial population size as well as the initial frequency of cooperators. Suppose there are n_0 individuals and a portion p_0 of them are cooperators initially. Given Eq. 11, the fixation probability $\rho(p_0, n_0)$ of cooperators satisfies the following backward Kolmogorov equation (in the following derivation, we omit the subscripts and write the fixation probability as $\rho(p, n)$):

$$\begin{cases} \mathcal{G}\rho(p, n) = 0, \\ \rho(1, n) = 1, & n > 0 \\ \rho(0, n) = 0, & n > 0 \end{cases} \quad [63]$$

where

$$\begin{aligned} \mathcal{G}f = & \left[-scp(1-p) + \delta_2 sc \frac{p(1-p)}{n} \right] \frac{\partial f}{\partial p} + [\alpha + s(b-c)pn - \lambda n^2] \frac{\partial f}{\partial n} \\ & + \frac{(\delta_1 B + D + \lambda n)p(1-p)}{2n} \frac{\partial^2 f}{\partial p^2} + \frac{(\delta_1 B + D + \lambda n)n}{2} \frac{\partial^2 f}{\partial n^2} \end{aligned} \quad [64a]$$

is the infinitesimal generator of Eq. 11. In the neutral case ($s = 0$), we can verify that the solution of Eq. 63 is $\rho = p$. Thus, the fixation probability equals the initial frequency p_0 when $s = 0$. For weak selection, inspired by the form of fixation probability in Eq. 20 and the boundary conditions of Eq. 63, we make the ansatz that the solution has the following form (5)

$$\rho(p, n) = p + sp(1-p)\phi(n) + o(s). \quad [65]$$

Substituting Eq. 65 into Eq. 63 and only retaining the first-order term of s (note that $\alpha = O(s)$ and $\lambda \ll s$ due to the large carrying capacity), we can obtain

$$-c + \frac{\delta_2 c}{n} - \frac{\delta_1 B + D}{n} \phi + \frac{(\delta_1 B + D)n}{2} \phi'' = 0. \quad [66]$$

This is an Euler-Cauchy equation, whose solution is

$$\phi(n) = A_1 n^2 + \frac{A_2}{n} - \frac{cn}{\delta_1 B + D} + \frac{\delta_2 c}{\delta_1 B + D}. \quad [67]$$

where A_1 and A_2 are arbitrary constants. Substitution into Eq. 65 yields

$$\rho(p, n) = p + sp(1-p) \left(A_1 n^2 + \frac{A_2}{n} - \frac{cn}{\delta_1 B + D} + \frac{\delta_2 c}{\delta_1 B + D} \right) \quad [68]$$

To satisfy the boundary conditions (i.e. $\rho(0, n) = 0$ and $\rho(1, n) = 1$), $p(1-p)\phi(n)$ must equal to 0 when $p = 0$ and $p = 1$, which has already been satisfied. Furthermore, when $c = 0$, cooperators and defectors always have the same payoffs. Thus, in this case, the evolution is also equivalent to a neutral drift and the fixation probability must equal p , which yields $p(1-p)\phi(n) = 0$ for all p when $c = 0$. We obtain $A_1 = A_2 = 0$. So the fixation probability of cooperators is

$$\rho = p_0 - \frac{sc}{\delta_1 B + D} p_0 (1-p_0) (n_0 - \delta_2). \quad [69]$$

Note that this expression for the fixation probability is similar to Eq. 20 (ignoring the difference in the coefficient), but the carrying capacity M has been replaced by initial population size n_0 . If

$$\delta_2 > n_0, \quad [70]$$

then the fixation probability exceeds p_0 , which means cooperation is favored by selection. The form of this condition is similar to Eq. 21. Moreover, numerical simulations and individual-based simulations (based on compound Poisson processes) both verify that this condition accurately predicts when cooperation will be favored in a slow-growing population (Fig. S6 and S7).

B.2. Extinction and persistence. When $\alpha \sim s$ the population size grows slowly. For small initial population size n_0 , fixation occurs rapidly, so that fixation typically occurs before the population reaches carrying capacity. After cooperators or defectors become fixed, the system then becomes a one-dimensional process, where all individuals have the same payoff and thus have the same variance in offspring number. The population will then either go extinct, or grow logistically until the population reaches the carry capacity (Fig. S8) (and thereafter persist for exponentially long in the carrying capacity, see Appendix B). Here, we analyze the process of extinction or persistence in the case when cooperators fix (the procedures are analogous in case of defectors fixing first).

If cooperators become fixed, the system then contains only cooperators for all subsequent evolution. Therefore it becomes a one-dimensional diffusion process, which is given by

$$dx = x[\alpha + s(b-c) - \lambda x]dt + \sqrt{x(\delta_1 B + D + \lambda x)}dW_t^{(1)}. \quad [71]$$

By analyzing the corresponding deterministic system ($\delta_1 = \delta_2 = 0$), the carrying capacity of cooperators is given by

$$M_C = \frac{\alpha + s(b - c)}{\lambda}. \quad [72]$$

For the stochastic system, we first analyze the case of no competition for resources (i.e., $\lambda = 0$, and so no carrying capacity). The equation becomes

$$dx = x[\alpha + s(b - c)]dt + \sqrt{x(\delta_1 B + D)}dW_t^{(1)}. \quad [73]$$

When we ignore the stochastic term, this equation represents exponential growth. The population size grows without bound. When the stochastic term is considered, given there are x_f cooperators after they become fixed, there are two scenarios: grow to infinity or go extinct, since $x = 0$ is an absorbing state. The probability that cooperators go extinct given that there are x_f cooperators after their fixation can be computed by (3)

$$P_C = \frac{\int_{x_f}^{\infty} s(z)dz}{\int_0^{\infty} s(z)dz}. \quad [74]$$

where

$$s(z) = \exp\left(\int -\frac{2E(z)}{D(z)}dz\right), E(z) = z[\alpha + s(b - c)], D(z) = \delta_1 Bz + Dz. \quad [75]$$

After some manipulations, we obtain

$$P_C = \exp\left(-\frac{2(s(b - c) + \alpha)x_f}{\delta_1 B + D}\right), \quad [76]$$

which is termed as perishing probability here. And the probability of growing to infinity is thus $1 - P_C$.

When competition for resources is considered ($\lambda \neq 0$), the population can never grow without bound. There are only one absorbing state $x = 0$ and the population size is bounded due to the restriction of carrying capacity. Thus, for all x_f , the probability that cooperators finally perish is 1. However, the population usually does not perish directly after they fix, but the population still grows to the carrying capacity and fluctuates around it for a very long time (in fact, exponentially long in the carrying capacity (see (7) and Appendix B)) before their extinction (Fig. S8). Thus, when the carrying capacity is large, in any practical application the population will not go extinct once it has reached carrying capacity. So we can artificially divide the dynamics into two scenarios: given there are x_f cooperators after they become fixed, the cooperators will either perish directly (Fig. S8A), or they grow to carrying capacity and then fluctuate around it for an extremely long time (Fig. S8C). Since the initial population size is small, the competition among individuals can be ignored before the population size reaches the carrying capacity. The probability of perishing directly can be approximated by the case of no competition, i.e. P_C . And the probability of reaching carrying capacity is approximated by $1 - P_C$. This approximation agrees well with simulations (Fig. S9A).

Similarly, if defectors become fixed, then the extinction probability can also be obtained using the same method. That is

$$P_D = \exp\left(-\frac{y_f \alpha}{\delta_1 B + D}\right). \quad [77]$$

And if they do not go extinct rapidly they will grow and finally reach their carrying capacity, which is given by

$$M_D = \frac{\alpha}{\lambda}. \quad [78]$$

Note that the carrying capacities for cooperators and defectors are no longer identical compared to the case of $\alpha \gg s$. The cooperators have an extra advantage in that their carrying capacity is larger than defectors'.

Now, we consider the two processes (fixation and extinction/persistence) together. Suppose there are x_0 cooperators and y_0 defectors in the population initially. We call the probability that cooperators become fixed and then keep growing to carrying capacity M_C the "persistence probability" (for cooperators). Thus, the persistence probability is determined by the combination of the fixation probability and (one minus) the extinction probability. However, the exact number of cooperators or defectors when they become fixed (i.e. x_f and y_f) is difficult to analyze. So it is difficult to give an explicit expression for the persistence probability. However, intuitively, since fixation is fast and the growth rate α is small, the population size after fixation will not change much compared to the initial population size. Thus, we expect that x_f and y_f have the same order as n_0 .

We have already shown that when $\delta_2 > M$, cooperation will be favored by the selection. Here, we show that even when $\delta_2 < M$, cooperators may have a greater persistence advantages than defectors. By Eq. 69, we see that when $\delta_2 < M$, larger variance in due to baseline births δ_1 is beneficial to the fixation of cooperators, but it also leads to more likely extinction after they fix (see Eq. 76). Thus we predict that intermediate δ_1 is most beneficial for cooperators' persistence. For defectors, larger δ_1 is detrimental to their fixation and also detrimental to their growth to M_D after fixation. Thus, large δ_1 is always detrimental to defectors persisting (Fig. S9B). For some intermediate δ_1 , cooperators may have greater persistence probability than defectors, which is a different sense in which offspring variance can foster the evolution of cooperation.

Similarly, we find that given a fixed initial frequency p_0 , larger initial population size n_0 is detrimental to cooperators' fixation but beneficial for their growth to carrying capacity M_C after fixation (since $x_f \sim n_0$). And so we also predict that an intermediate value of the initial population size n_0 is most beneficial to cooperators' persistence. But for defectors, larger n_0 is always beneficial for defector's persistence (Fig. S9C).

Based on the discussions above, we conclude that cooperators may have an advantage over defectors in stochastic populations for three distinct reasons: (1) larger fixation probability; (2) larger carrying capacity; (3) larger persistence probability.

Appendix A: Details of simulations

Two methods of simulation are used in this work: numerical simulations by sampling the SDE, and Monte Carlo simulations based on compound Poisson process. Here, we provide more details about these simulations.

Numerical simulations:

Numerical simulations are based on Eq. 8. We choose a small time step $\Delta t = 0.01$ and set the initial configuration (x_0 and y_0). The Euler-Maruyama scheme is used to sample Eq. 8 (8), which is given by

$$x_{t+\Delta t} = x_t + x_t(\alpha + s\pi_C - \lambda(x_t + y_t))\Delta t + \sqrt{x_t(\delta_1 B + \delta_2 s\pi_C + D + \lambda(x_t + y_t))}\Delta W_t^{(1)} \quad [79a]$$

$$y_{t+\Delta t} = y_t + y_t(\alpha + s\pi_D - \lambda(x_t + y_t))\Delta t + \sqrt{y_t(\delta_1 B + \delta_2 s\pi_D + D + \lambda(x_t + y_t))}\Delta W_t^{(2)}. \quad [79b]$$

Here, $\Delta W_t^{(1)} \sim N(0, \Delta t)$ and $\Delta W_t^{(2)} \sim N(0, \Delta t)$ are two independent Gaussian random variables. In each time step, by sampling random numbers from $\Delta W_t^{(1)}$ and $\Delta W_t^{(2)}$, we can obtain the population composition in the next time step. By iteration of this procedure, a trajectory of evolution is obtained. Then, we can simulate a large number replicate trajectories and compute the fixation probability.

Monte Carlo simulations:

Monte Carlo simulations must rely on a explicit reproducing process. Here, as we showed in Section 3, we assume the birth process is a compound Poisson process and the death process is a classic Poisson process.

For compound Poisson process, the times of birth events is still a classic Poisson process. Thus, we can consider the birth events and death events together. We refer to them collectively as the updating events. In the following, we only illustrate the procedure of simulation by taking the Poisson-Poisson process as an example, which is similar to Poisson-Negative binomial process.

For a cooperator, its birth event takes place at rate θ_C (i.e., the parameter of M_t in Section 3), and θ_D for a defector. In each birth event, the litter size obeys a Poisson distribution with parameter μ_C (μ_D for a defector). Using the property of Poisson process that the sum of two Poisson process is also a Poisson process with rate summed, the birth events of all cooperators take place at rate $x\theta_C$, and $y\theta_D$ for defectors. Similarly, the death events take place at rate $Dx + \lambda x(x + y)$ for cooperators, and $Dy + \lambda y(x + y)$ for defectors. Thus, the updating events take places with rate

$$R = x\theta_C + y\theta_D + Dx + \lambda x(x + y) + Dy + \lambda y(x + y). \quad [80]$$

In the simulation, we first set the initial configuration of population (x_0 and y_0). Suppose in time T , there are x_T cooperators and y_T defectors. Since the time interval between two consecutive events of a Poisson process follows an exponential distribution with the same parameter of the Poisson process, we sample a value of ΔT according to an exponential distribution with parameter R . Then, in time $T + \Delta T$, the population composition is computed by the following rule:

Step 1. We generate a random number r which is uniformly sampled from $[0, 1]$.

Step 2. The population composition updates in one of the following ways according to the value of r .

- Cooperators' birth: If $r < x\theta_C/R$ (with probability $x\theta_C/R$), then $x_{T+\Delta T} = x_T + \delta x$ and $y_{T+\Delta T} = y_T$, where δx is an integer randomly sampled from a Poisson distribution with parameter μ_C (Poisson-Poisson process).
- Defectors' birth: If $x\theta_C/R \leq r < (x\theta_C + y\theta_D)/R$ (with probability $y\theta_D/R$), then $x_{T+\Delta T} = x_T$ and $y_{T+\Delta T} = y_T + \delta y$, where δy is an integer randomly sampled from a Poisson distribution with parameter μ_D (Poisson-Poisson process).
- Cooperators' death: If $(x\theta_C + y\theta_D)/R \leq r < (x\theta_C + y\theta_D + Dx + \lambda x(x + y))/R$ (with probability $(Dx + \lambda x(x + y))/R$), then $x_{T+\Delta T} = x_T - 1$ and $y_{T+\Delta T} = y_T$.
- Defectors' death: If $((x\theta_C + y\theta_D + Dx + \lambda x(x + y))/R \leq r < 1$ (with probability $(Dy + \lambda y(x + y))/R$), then $x_{T+\Delta T} = x_T$ and $y_{T+\Delta T} = y_T - 1$.

Thus, we obtain the population composition in time $T + \Delta T$. Next, we sample another ΔT , and the algorithm enters the next cycle. Recording all time steps and the corresponding population composition, we obtain a trajectory of evolution.

Appendix B: Proof for exponential extinction time

Here, we show that if the number of a phenotype has reached its carrying capacity, the expected time for its extinction is exponentially long in the carrying capacity.

We consider a phenotype whose growth rate is r once fixed in the population ($r = \alpha + s(b - c)$ for cooperators and $r = \alpha$ for defectors). Supposing the carrying capacity is M , we have $\lambda = r/M$. The evolutionary equation after they are fixed becomes

$$dx = rx \left(1 - \frac{x}{M}\right) dt + \sqrt{x(\delta_1 B + D)} dW_t^{(1)}. \quad [81]$$

Supposing the population has already reached the carrying capacity ($x = M$), the expected time of extinction (denoted by τ) is (3)

$$\mathbb{E}[\tau] = 2 \int_0^M \left(\int_0^z s(y) dy \right) m(z) dz \quad [82]$$

where

$$s(z) = \exp \left(\int -\frac{2rz(1 - z/M)}{z(\delta_1 B + D)} dz \right) = \exp \left(-\frac{2r}{\delta_1 B + D} \left(z - \frac{z^2}{2M} \right) \right), \quad [83a]$$

$$m(z) = 1/(z(\delta_1 B + D)s(z)). \quad [83b]$$

Remembering that for a convex function $f(x)$, it satisfies

$$(b - a)f \left(\frac{a + b}{2} \right) < \int_a^b f(x) dx, \quad [84]$$

we have

$$\begin{aligned} \mathbb{E}[\tau] &= 2 \int_0^M \left(\int_0^z s(y) dy \right) m(z) dz \\ &> 2 \int_0^M z \exp \left(-\frac{2r}{\delta_1 B + D} \left(\frac{z}{2} - \frac{z^2}{8M} \right) \right) m(z) dz && \text{[since } s(x) \text{ is convex]} \\ &= \frac{2}{\delta_1 B + D} \int_0^M \exp \left(-\frac{2r}{\delta_1 B + D} \left(-\frac{z}{2} + \frac{3z^2}{8M} \right) \right) dz \\ &= \frac{2}{\delta_1 B + D} \int_0^M \exp \left(-\frac{2r}{\delta_1 B + D} \frac{z}{2} \left(-1 + \frac{3z}{4M} \right) \right) dz \\ &> \frac{2}{\delta_1 B + D} \int_0^M \exp \left(\frac{r}{\delta_1 B + D} \frac{z}{4} \right) dz && \text{[since } z(1 - \frac{3z}{4M}) > \frac{z}{4}] \\ &= \frac{8}{r} \exp \left(\frac{r}{4(\delta_1 B + D)} M \right). \end{aligned} \quad [85]$$

Thus, for sufficiently large M , the time for extinction is very long. In practice, this duration of time may be so long that we reasonably say the population will persist, conditional on reaching carrying capacity.

Supplementary figures

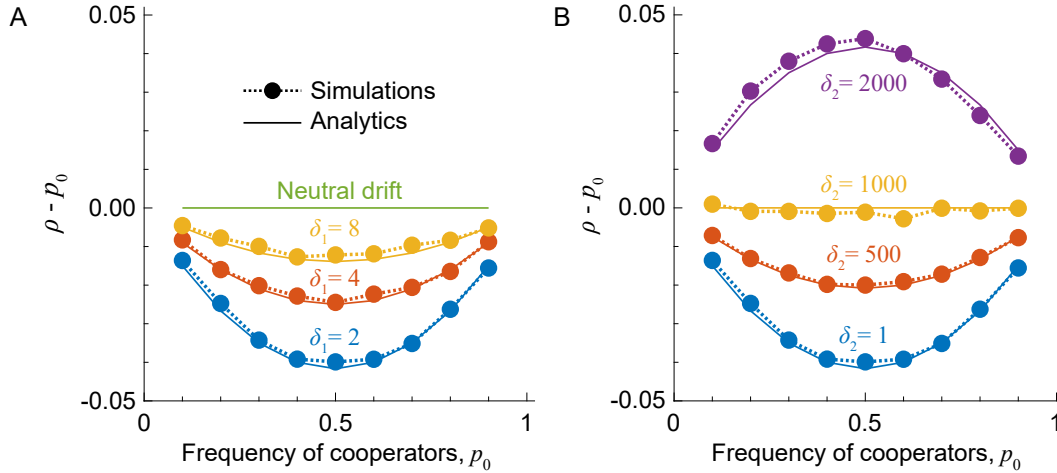


Fig. S1. The initial frequency of cooperators and δ_1 can affect the fixation probability. Under neutral drift, the fixation probability equals to the initial frequency of cooperators, p_0 . The panels show the fixation probabilities minus p_0 . If $\rho - p_0$ exceeds zero cooperation is favored by natural selection, and vice versa. (A) For all δ_1 , cooperation is never favored by selection, although large δ_1 makes the fixation probability closer to neutral drift. Furthermore, the fixation probability for intermediate initial frequency p_0 deviates the most from neutrality. (B) By contrast, large δ_2 can cause selection to favor fixation of cooperators. For the case of $M = 1,000$, when $\delta_2 < 1,000$, cooperation is never favored and $p = 0.5$ is the most detrimental to the evolution of cooperation. But when $\delta_2 > M$, all fixation probabilities exceed that of neutral drift, and $p_0 = 0.5$ is the most beneficial to cooperation. And when $\delta_2 = M$, the fixation probability equals to the neutral case. Numerical simulations (dots) agree well with analytical approximations (dashed lines). Parameters: $B = 2, D = 1, b = 1.1, c = 1, s = 0.001, n_0 = 500, \lambda = 10^{-3}$ ($M = 1000$), $\delta_2 = 1$ (A), $\delta_1 = 2$ (B).

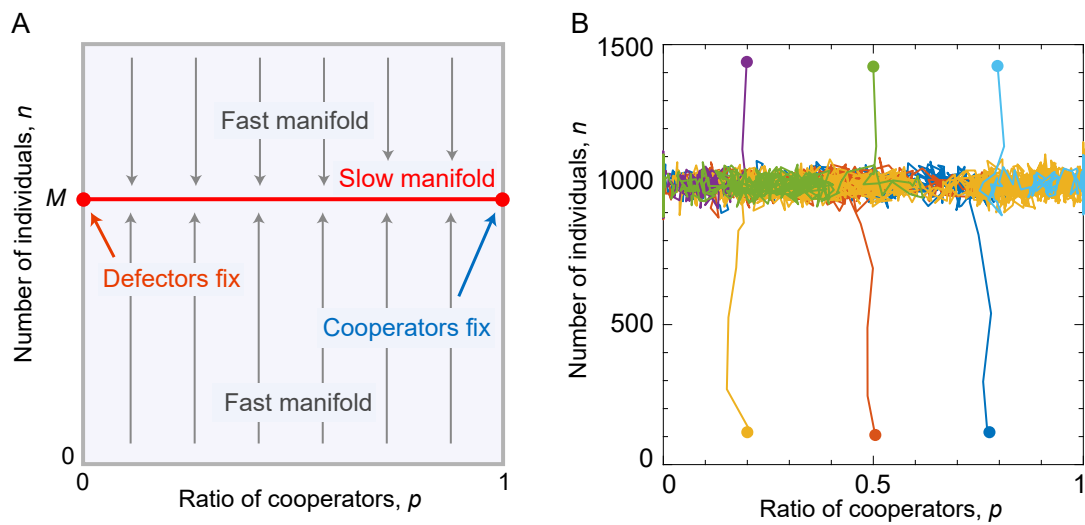


Fig. S2. Simulations showing fast and slow manifolds. (A) When the growth rate is much larger than selection intensity ($\alpha \gg s$), population size equilibrates much faster than the frequency of cooperators: n is the fast variable and p is the slow variable. In a deterministic analysis, all trajectories rapidly approach the slow manifold, without changes in p , and then move along the slow manifold. (B) Simulations of the stochastic model illustrate the timescale separation. For six different settings of initial states (solid point), we show trajectories produced by numerical simulation of Eq. 8. The trajectories exhibit behavior that agrees well with the deterministic analysis of the fast-slow system. Parameters: $B = 2$, $D = 1$, $\lambda = 0.001$, $\delta_1 = \delta_2 = 1$, $b = 1.1$, $c = 1$, $s = 0.001$.

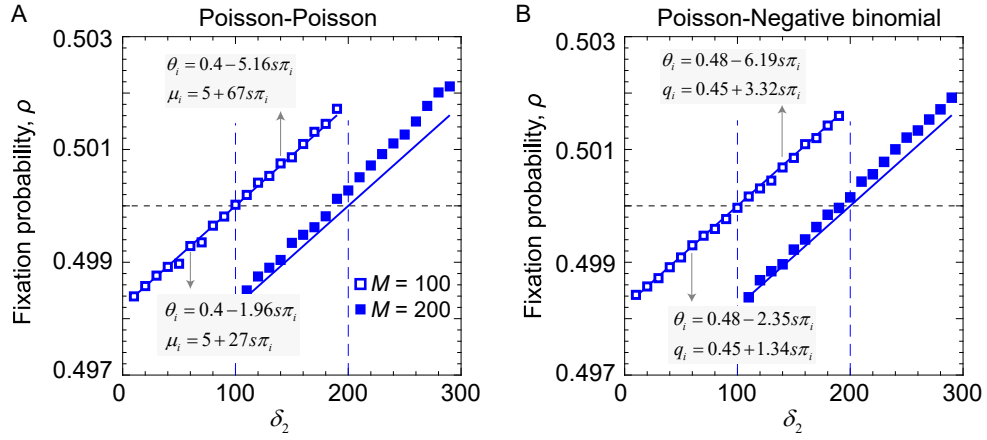


Fig. S3. Selection for cooperation in a compound Poisson birth process. We simulated a compound Poisson birth process with either a Poisson-distributed litter size (A) or a negative binomial litter size (B). The parameters of the birth process (θ_i and μ_i in panel A; θ_i and q_i in panel B) can be chosen to satisfy our general conditions for the mean and variance in total offspring produced per unit time, for any choice of $\delta_1 > 1$, δ_2 , and B . Two examples with the parameters that correspond to $(\delta_1 = 6, \delta_2 = 60)$ and $(\delta_1 = 6, \delta_2 = 140)$ are shown in each panel. Blue squares indicate the fixation probability of cooperators, starting from an initial population with $x_0 = y_0 = 50$, observed in 5×10^7 replicate Monte Carlo simulations, with carrying capacity either $M = 100$ or $M = 200$. Selection favors cooperation if the fixation probability ρ exceeds the initial fraction of cooperators, 0.5 (horizontal dashed line). The solid lines plot our analytical approximation for the fixation probability (Eq. 20). As predicted by our analysis, cooperation is favored when $\delta_2 > M$. Parameters: $B = 2$, $D = 1$, $\delta_1 = 6$, $s = 0.001$, $b = 1.1$, $c = 1$, $m = 5$ (negative binomial), $x_0 = y_0 = 50$, $\lambda = 1/100$ ($M = 100$) or $\lambda = 1/200$ ($M = 200$).

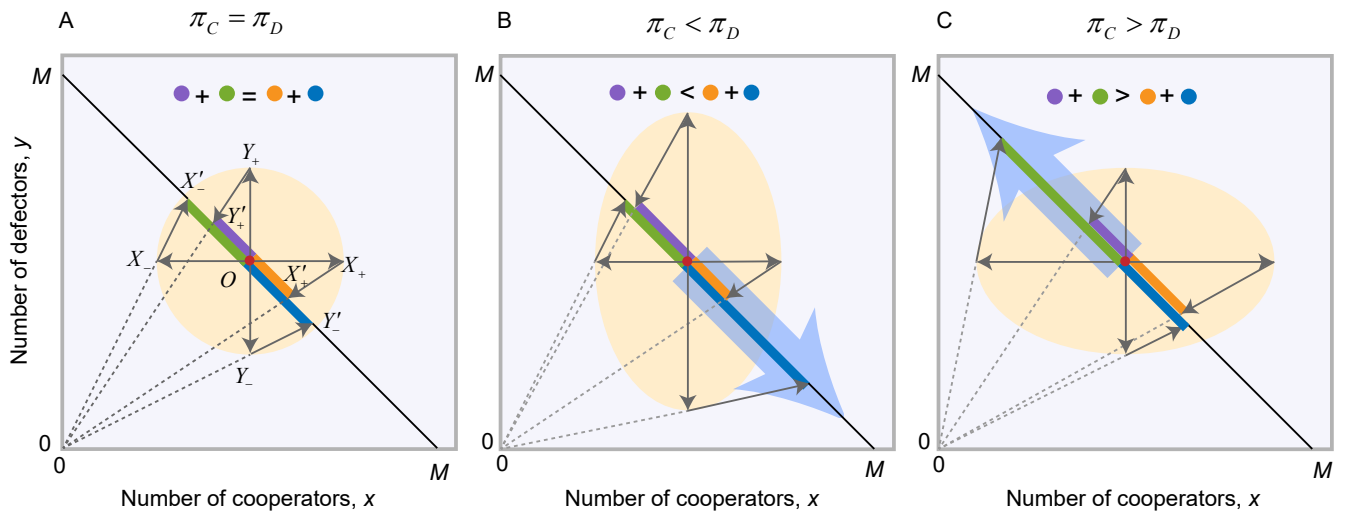


Fig. S4. How demographic stochasticity can favor cooperation or defection. Stochastic fluctuations away from the slow manifold, followed by rapid return, can induce an advective force on the frequency of cooperators. For simplicity we consider constant payoffs, where π_C and π_D are independent of the number of cooperators and defectors. The ellipses illustrate the variance-covariance structure of two-dimensional Gaussian fluctuations around the slow manifold from a given point $x = M/2$ and $y = M/2$ (red point O). (A) When $\pi_C = \pi_D$, fluctuations from point O are isotropic, shown as a circle. We consider four representative fluctuations from point O , X_- , X_+ , Y_- , Y_+ , and the following points of return X'_- , X'_+ , Y'_- , Y'_+ to the slow manifold. For isotropic fluctuations there is no expected change in p after return to the slow manifold. (B) For $\pi_C < \pi_D$, the Gaussian fluctuations are an-isotropic, shown as an ellipse, with larger fluctuations in the number of defectors. This asymmetry leads to an expected increase in cooperator frequency p after return to the slow manifold, as indicated by the blue arrow. (C) For $\pi_C > \pi_D$, the larger fluctuation occurs in the number of cooperators, which leads to an expected decrease in cooperator frequency after return to the slow manifold. These effects of an-isotropic noise are similar to those discussed by (9), but they arise here even when both types have the same baseline birth rate and the same carrying capacity.

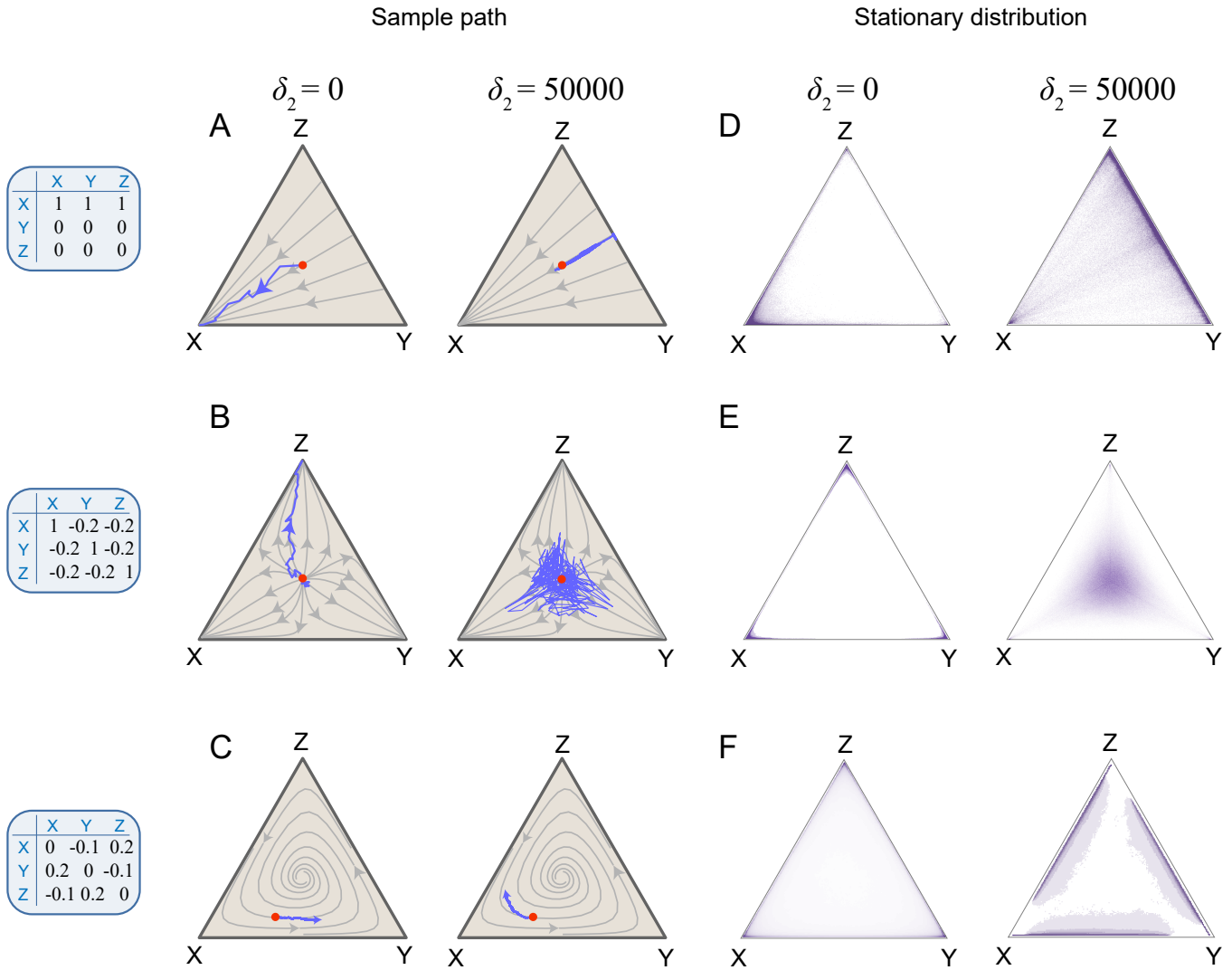


Fig. S5. Demographic noise produces qualitatively different evolutionary outcomes in a game with three strategies. We show evolutionary trajectories in simplices representing three strategic frequencies in three different games (rows A, B, C). Blue trajectories show simulations of the stochastic system starting from the red point; and gray trajectories show the dynamics predicted by the classic replicator dynamics in an infinite population. When the offspring variance related to payoff, δ_2 , is sufficiently large, it can convert an unstable (stable) equilibrium point into a stable (unstable) equilibrium (A, B). For the rock-paper-scissors game (row C), the classic replicator dynamics predicts a spiral sink towards a stable point with all three strategies present. The stochastic system does not follow any of these trajectories exactly, but its direction can be reversed when δ_2 is large (panel C shows the ensemble average trajectory over 1000 simulations). Furthermore, if we assume that the boundary is reflecting, the stationary distributions (conditioned on $n > 0$) also show that large δ_2 qualitatively changes the qualitatively (D, E, F). Parameters: $B = 2$, $D = 1$, $s = 0.001$, $\delta_1 = 2$, $\lambda = 10^{-4}$.

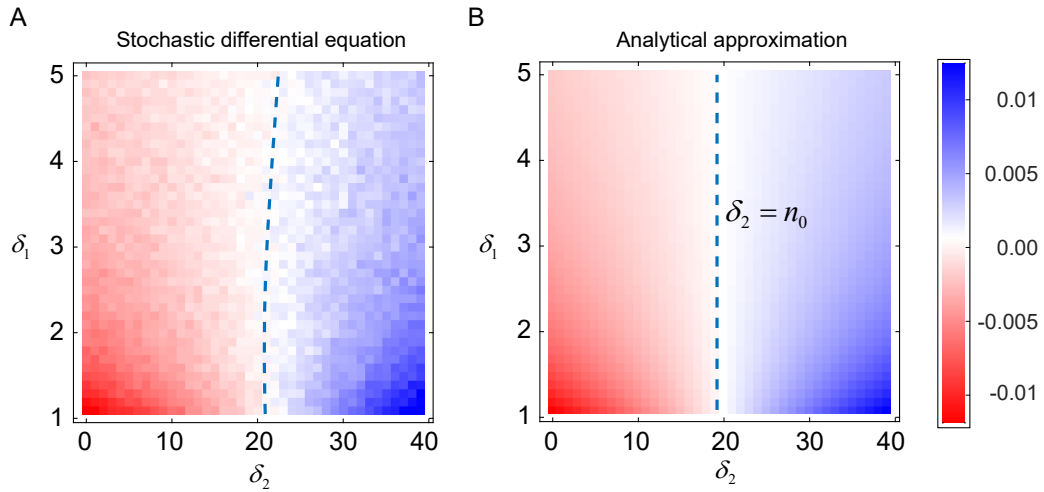


Fig. S6. Cooperation is favored if $\delta_2 > n_0$ in a slow-growing population. When the baseline growth rate is small, namely $\alpha \sim s$, we can still obtain an analytic prediction for the fixation probability of cooperation. In this figure, each pixel represents the fixation probability minus p_0 (the neutral fixation probability). The blue region means cooperation is favored and the red region means defection is favored. (A) shows the fixation probabilities sampled from the stochastic differential equation (Eq. 8). (B) is obtained from our analytical approximation (Eq. 69). As predicted, when δ_2 exceeds the initial population size, namely, $\delta_2 > n_0$, cooperators are favored by selection in a slow-growing population. Parameters: $b = 1.1$, $c = 1$, $s = 0.005$, $\lambda = 10^{-5}$, $B = 2$, $D = 1.995$.

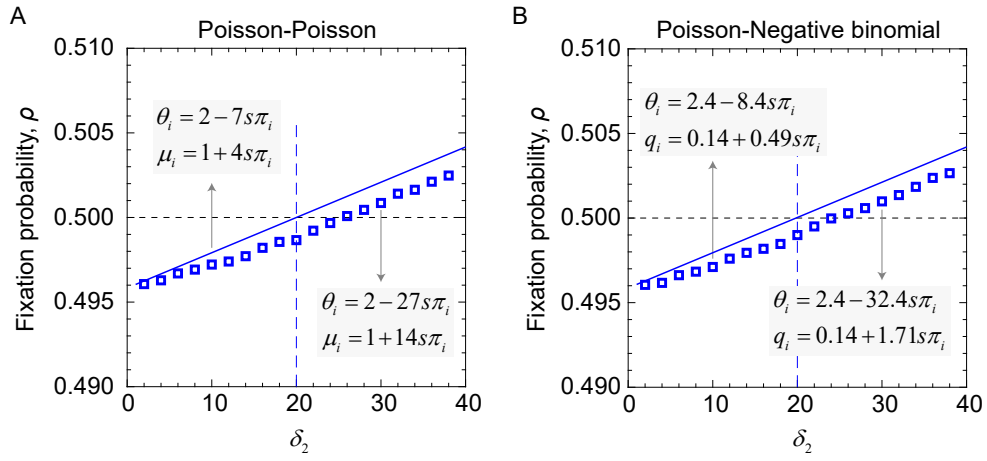


Fig. S7. Selection for cooperation in a compound Poisson process, for a slow-growing population. Here, we consider the case of a slow-growing population, with $\alpha \sim s$. We simulated a compound Poisson birth process with either a Poisson-distributed litter size (A) or a negative binomial litter size (B). The parameters of the birth process (θ_i and μ_i in panel A; θ_i and q_i in panel B) can be chosen to satisfy our general conditions for the mean and variance in total offspring produced per unit time, for any choice of $\delta_1 > 1$, δ_2 , and B . Two examples with the parameters that correspond to $(\delta_1 = 2, \delta_2 = 10)$ and $(\delta_1 = 2, \delta_2 = 30)$ are shown in each panel. Blue squares indicate the fixation probability of cooperators, starting from an initial population with $x_0 = y_0 = 10$ ($n_0 = 20$), observed in 10^7 replicate Monte Carlo simulations. Selection favors cooperation if the fixation probability ρ exceeds the initial fraction of cooperators, $p_0 = 0.5$ (horizontal dashed line). The solid lines plot our analytical approximation for the fixation probability (Eq. 69). As predicted by our analysis, cooperation is favored when $\delta_2 > n_0$ in a slow-growing population. Parameters: $B = 2$, $D = 1.998$, $s = 0.005$, $b = 1.1$, $c = 1$, $\lambda = 10^{-6}$, $\delta_1 = 2$, $m = 5(B)$.

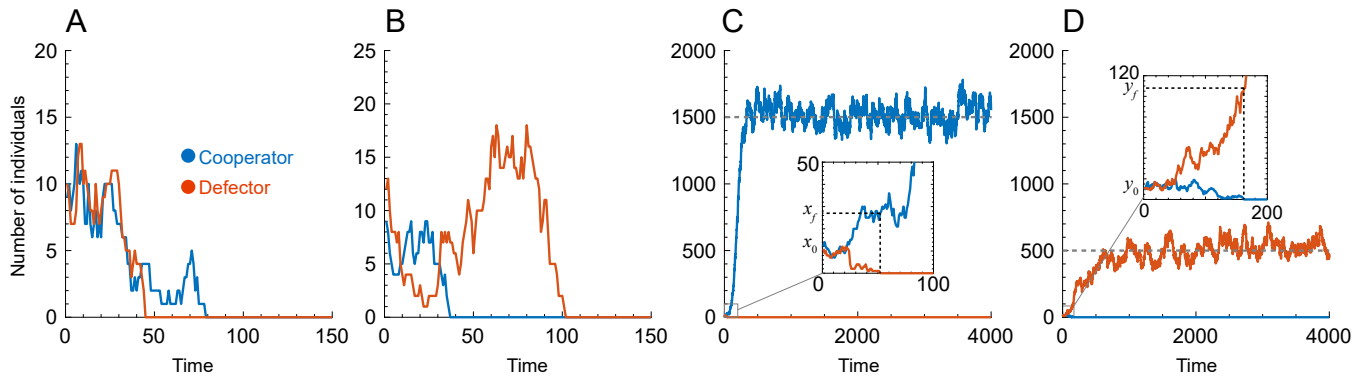


Fig. S8. Sample trajectories of evolution in a slow-growing population. If the baseline growth rate is small ($\alpha \sim s$), given a small initial population then fixation always occurs before the population size reaches carrying capacity. Since there is only one absorbing state ($x = y = 0$), all trajectories eventually absorb into extinction. However, if one phenotype reaches carrying capacity, the time to absorption is exponentially long in the carrying capacity, so that the population effectively persists for large finite times. We therefore classify the trajectories into four types: (A) Cooperators fix but then perish. (B) Defectors fix but then perish. (C) Cooperators fix and then reach their carrying capacity M_C . (D) Defectors fix and then reach their carrying capacity M_D . Parameters: $b = 3$, $c = 1$, $s = 0.01$, $B = 0.25$, $D = 0.24$, $x(0) = y(0) = 10$, $\lambda = 2 \times 10^{-5}$, $\delta_1 = \delta_2 = 1$.

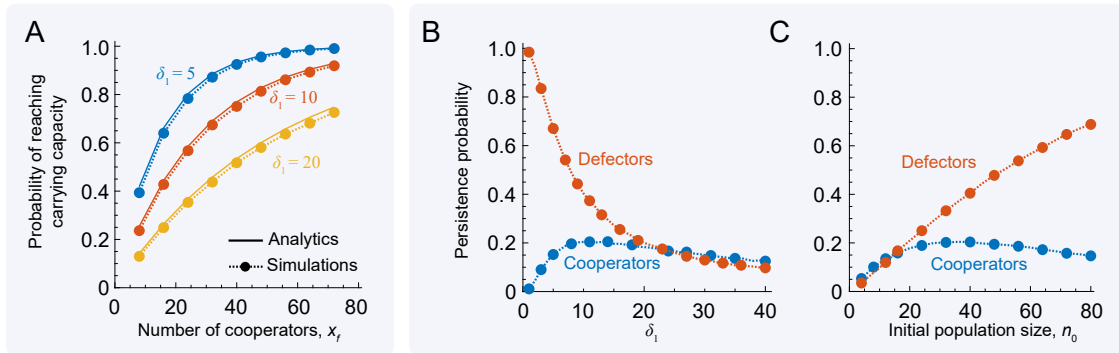


Fig. S9. Persistence probabilities of cooperators and defectors in slow-growing populations. When the baseline growth rate is small (i.e., $\alpha \sim s$), fixation of one type or the other will occur before the population reaches carrying capacity. The probability that a phenotype not only fixes, but also reaches its carrying capacity in subsequent evolution is called the persistence probability. (A) Take cooperators as an example. If there are x_f cooperators when they become fixed, the probability that they grow to carrying capacity can be estimated by the probability that the population grows without bound in the case without resource competition. This approximation agrees well with simulations. (B) For cooperators, values of δ_1 that are either too large or too small are both detrimental to cooperator persistence. Intermediate values of δ_1 are most beneficial to cooperator persistence. For defectors, larger δ_1 is always detrimental to defector persistence. (C) Similarly, Intermediate values of the initial population size is most beneficial to cooperator persistence. However, larger initial population sizes are always beneficial for defector persistence. These results agree with our theoretical analysis in Section B.2. Parameters: $b = 2$, $c = 1$, $s = 0.01$, $B = 0.1$, $D = 0.09$, $\lambda = 10^{-5}$, $\delta_2 = 1$, $\delta_1 = 10$ (C), $x_0 = y_0 = 20$ (B), $p_0 = 0.5$ (C).

References

1. A Lindén, S Mäntyniemi, Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology* **92**, 1414–1421 (2011).
2. JM Ver Hoef, PL Boveng, Quasi-poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology* **88**, 2766–2772 (2007).
3. CW Gardiner, *Handbook of stochastic methods*. (Springer, Berlin), (2004).
4. TL Parsons, T Rogers, Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: A constructive approach with examples from theoretical biology. *J. Phys. A Math. Theor.* **50**, 415601 (2017).
5. P Czuppon, A Traulsen, Fixation probabilities in populations under demographic fluctuations. *J. Math. Biol.* **77**, 1233–1277 (2018).
6. A McAvoy, B Allen, Fixation probabilities in evolutionary dynamics under weak selection. *J. Math. Biol.* **82**, 14 (2021).
7. A McAvoy, N Fraiman, C Hauert, J Wakeley, MA Nowak, Public goods games in populations with fluctuating size. *Theor. Popul. Biol.* **121**, 72–84 (2018).
8. E Platen, An introduction to numerical methods for stochastic differential equations. *Acta Numer.* **8**, 197–246 (1999).
9. GW Constable, T Rogers, AJ McKane, CE Tarnita, Demographic noise can reverse the direction of deterministic selection. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E4745–E4754 (2016).