PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Social evolution under demographic stochasticity

David V. McLeod[1]*, **Troy Day**[2]

1 Institute for Integrative Biology, ETH Zürich, Zürich, Switzerland, **2** Department of Mathematics and Statistics, Department of Biology Queen's University, Kingston, ON, Canada

* david.mcleod@env.ethz.ch

Abstract

How social traits such as altruism and spite evolve remains an open question in evolutionary biology. One factor thought to be potentially important is demographic stochasticity. Here we provide a general theoretical analysis of the role of demographic stochasticity in social evolution. We show that the evolutionary impact of stochasticity depends on how the social action alters the recipient's life cycle. If the action alters the recipient's death rate, then demographic stochasticity always favours altruism and disfavours spite. On the other hand, if the action alters the recipient's birth rate, then stochasticity can either favour or disfavour both altruism and spite depending on the ratio of the rate of population turnover to the population size. Finally, we also show that this ratio is critical to determining if demographic stochasticity can reverse the direction of selection upon social traits. Our analysis thus provides a general understanding of the role of demographic stochasticity in social evolution.

Author summary

Explaining the evolution of social traits such as altruism and spite remains a key outstanding problem in evolutionary biology. Here we develop a simple theory for the effect of demographic stochasticity (random variation in an individual's birth and death rates) on the evolution of social traits. Our results provide a clear set of predictions: whether a social trait is favoured or disfavoured is determined by how the social action alters the recipient's life cycle. If the social action alters the recipient's death rate, then altruism is favoured and spite disfavoured. If instead the social action alters the recipient's birth rate, then both altruism and spite can be either favoured or disfavoured—the precise outcome depends upon the ratio of the population turnover rate to the population size.

OPEN ACCESS

Citation: McLeod DV, Day T (2019) Social evolution under demographic stochasticity. PLoS Comput Biol 15(2): e1006739. https://doi.org/10.1371/journal.pcbi.1006739

Editor: Joshua B. Plotkin, University of Pennsylvania, UNITED STATES

Received: July 12, 2018

Accepted: December 23, 2018

Published: February 4, 2019

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The evolution of social traits remains a very active area of investigation in evolutionary biology [1–4]. Research has predominately focused upon how different mechanisms such as population structure [5–8], kin discrimination [9–11] or greenbeard effects [3, 12, 13] create

heterogeneity in interactions among individuals of different types, leading to the evolution of social traits. Recent evolutionary theory, however, has considered whether or not in the absence of interaction heterogeneity, demographic stochasticity alone can promote the evolution of altruism (e.g., [14–16]; see also [17] for when interaction heterogeneity and demographic stochasticity work in combination). These studies concluded that since altruism increases population size, it confers a stochastic advantage that can reverse (weak) selection against altruism. Counterexamples to this prediction have been found however (e.g., see below) leading one to wonder whether unambiguous conclusions can be drawn.

To address this question we develop a general theoretical analysis of the role of demographic stochasticity in social evolution for well-mixed populations. We start with a detailed description of birth and death events at the individual level [18] and then derive a very simple theory that makes a set of clear, general, predictions. Whether a social trait is favoured or disfavoured by demographic stochasticity is determined by how the social action alters the recipient's life cycle. When the action alters the recipient's death rate, altruism is stochastically favoured, and spite is stochastically disfavoured. When the action alters the recipient's birth rate both altruism and spite can be either stochastically favoured or disfavoured, with the outcome depending upon the ratio of the rate of population turnover to the population size. These results provide a general understanding of the role of demographic stochasticity in the evolution of social traits. They also explain previous models and counterexamples, and illustrate how previous results are special cases of a simple general principle.

Models

Consider a well-mixed population of size $\Omega n(t)$, where Ω is the habitat size and $n(t)$ is the population density at time t. The population consists of two types of individuals: type 1 individuals, who are social actors capable of altering the birth or death rate of other individuals in the population, and type 2 individuals who are not. The social action may occur through direct contact between individuals or by the production and uptake of an external compound (e.g., the release of siderophores or toxins by bacteria [19–21]). Each individual in the population is equally likely to be the recipient of the social action, and the effect of the social action upon the recipient is identical among types. We distinguish between two possible social traits: *altruism*, which we define to be an action that enhances the vital rates of other individuals (e.g., by increasing birth rates or decreasing mortality rates) and *spite*, which we define to be an action that inhibits the vital rates of other individuals (e.g., by decreasing birth rates or increasing mortality rates). These are standard definitions if the social trait comes at a cost to the actor [22]. Thus at demographic equilibrium the population size n will increase as the frequency of altruism increases whereas it will decrease as the frequency of spite increases.

Denote the per-capita birth and mortality rates as b and m, respectively, and let the per-capita cost of the social trait be ϵc, where ϵ is a parameter controlling the magnitude of the costs (b, m, and c may depend upon population densities and/or the state of the environment). Thus the per-capita growth rate of social actors (type 1 individuals) and non-actors (type 2 individuals) is $b - m - \epsilon c$ and $b - m$, respectively. As a consequence, whenever $\epsilon > 0$, non-actors have a selective advantage, and so in the absence of mutations and stochasticity they will ultimately take over the population. If $\epsilon = 0$, then the social trait is cost-free and so neither type of individual is selectively favoured. Finally, we suppose that mutation between the two types occurs at a per-capita rate μ. We will further assume that in the absence of selection and mutation, there is an asymptotically stable curve of ecological equilibria given by $b = m$. This is a curve rather than a fixed point because in the absence of selection and mutation, both types have identical per-capita growth rates.

If selection is weak, mutations rare, and habitat size large, then the system dynamics occur on two timescales: a fast timescale corresponding to demographic processes (birth and death events) and a slow timescale corresponding to evolutionary change in population composition. As our primary interest is the evolution of the population, our focus is on the slow timescale. On this slow timescale, let p be the fraction of social actors (type 1 individuals); then we can rewrite total population density as a function of $p(t)$ alone, that is, $n(t) = n(p(t)) = n(p)$ (see S1 Appendix). Then let $b(p)$, $m(p)$, and $c(p)$ be the per-capita birth, death, and costs on the slow timescale. If ϵ is small then $T(p) \equiv b(p) + m(p)$ is approximately the total rate at which demographic events are occurring and so is a measure of the rate of population turnover. Formally, it is also the variance in per-capita growth rate at selective neutrality.

Using a diffusion approximation of the full, individual-based, stochastic process [23, 24] (see S1 Appendix) and eliminating the fast timescale dynamics [25–28], the evolutionary change in frequency of social actors in the population is described by the stochastic differential equation (SDE)

$$dp = \alpha(p)dt + \sqrt{\sigma^2(p)}dW_t \tag{1}$$

where $\alpha(p) \equiv \mu(1-2p) - \epsilon c(p)p(1-p)$, $\sigma^2(p) \equiv p(1-p)T(p)/[\Omega n(p)]$, W_t is a Wiener process and we have neglected terms of order ϵ/Ω and μ/Ω (see S1 Appendix). Eq 1 is associated with a one-dimensional diffusion process with infinitesimal mean and variance $\alpha(p)$ and $\sigma^2(p)$ [29, 30]; when written as an SDE, the expression $\alpha(p)dt$ is often referred to as the "drift term". If mutation rate is sufficiently large, the diffusion process admits a stationary distribution, which we will denote by $\pi(p)$.

Note that in contrast to previous work (e.g., [16]), here our focus is the frequency of the social trait, p, rather than the density of social actors, $n(p)p$. As a consequence, there are no noise-induced effects in the drift term of Eq 1, whereas there are often noise-induced effects in the drift term of the SDE describing the change in density of social actors (see S1 Appendix, and also [16]). We opt to focus upon the frequency SDE rather then the density SDE because we are concerned with evolutionary processes, and evolution is a change in frequency not density.

Results

We wish to use Eq 1 to determine if stochasticity favours one type over another. Since $\alpha(p)dt$ represents the expected change in frequency of the social actors, while $\sqrt{\sigma^2(p)}dW_t$ represents stochastic noise around this mean change, one is tempted to simply examine the sign of $\alpha(p)$. With this approach, if $\alpha(p) < 0$ then the social actor (type 1) is disfavoured, which is the same conclusion as the deterministic model ($\Omega \to \infty$), and so this approach fails to take into account the role played by stochasticity. A second approach would be to suppose that whenever a mutation arises, it is either lost or sweeps to fixation before another mutation occurs, and so evolution proceeds according to a mutation-fixation process [31, 32]. With this approach, assessing if a trait is favoured or not is often done by comparing the probability a trait i mutant sweeps to fixation in a population monomorphic for trait j to the role-reversed situation (a comparison of invasion probabilities). If the costs of the social behaviour due to selection are sufficiently weak, $\epsilon \approx 0$, then from Eq 1 the invasion probability of a single social actor in a population of non-social individuals is $1/[\Omega n(0)]$, whereas the invasion probability of a single non-social individual in a population of social actors is $1/[\Omega n(1)]$. Hence a comparison of invasion probabilities favours the social actor whenever the social trait increases population size (altruism) [14–16]. The problem with comparing invasion probabilities alone is doing so fails to consider the full evolutionary process. Because in a mutation-fixation process

the population transitions from monomorphic state to monomorphic state, we can construct a Markov chain on the space of possible traits by letting N_i be the size of a trait-i population and μ_{ij} be the per-capita rate at which trait i mutates to trait j. Then the population will transition from a trait i state to a trait j state at a rate $\mu_{ij}N_i \times (1/N_i) = \mu_{ij}$. Thus in the absence of any biases in per-capita mutation rate, the population is equally likely to be observed in any monomorphic state, irrespective of the effect the trait has upon population size [32–34].

What both of these approaches have failed to take into account is the speed at which the change in population composition (and hence the evolutionary process) occurs. In particular, although the stochastic noise does not induce an average directionality to the change in p, the amount of stochasticity nevertheless is typically different for different values of p, and this will effect the speed at which the population composition changes, affecting the likelihood of observing the process in a particular state. As an analogy, a biased random walk whose step-size and time between steps depends upon the position of the walker will tend to spend more time in regions with smaller step-sizes and less frequent steps, independent of any bias in the directionality of the walk. Thus we will say that the social actor is favoured if, in the long-term, we are more likely to observe the system in a state in which the social actor is at greater frequency than the non-social actor (see S1 Appendix). For example, in the case where a stationary distribution $\pi(p)$ exists, the social actor is favoured if $\int_{1/2}^{1} \pi(p)dp > 1/2$.

To understand how this applies to the stochastic process defined by Eq 1, first suppose the social trait is cost-free ($\epsilon = 0$). Then the behaviour of Eq 1 is determined by two factors: the magnitude of the mutation rate μ and the ratio $T(p)/n(p)$. Mutation does not directly favour one type over the other and therefore the ratio $T(p)/n(p)$ should play a critical role in determining the values of p at which the system spends the most time. The following derivative tells us how this ratio changes with p:

$$\frac{d}{dp}\left[\frac{T(p)}{n(p)}\right] = \frac{T(p)}{n(p)}\left(\underbrace{-\frac{dn/dp}{n(p)}}_{\text{(i)}} + \underbrace{\frac{dT/dp}{T(p)}}_{\text{(ii)}}\right). \tag{2}$$

There are two components to Eq 2, each with a simple biological interpretation: (i) is the effect the social trait has upon population size, $n(p)$, and (ii) is the effect the social trait has upon population turnover, $T(p)$. In terms of our random walk analogy, as the population size increases, the step size of the random walk (in terms of frequency p) decreases, meaning that the process will tend to spend more time at values of p corresponding to large population sizes. Put another way, larger populations are more buffered against demographic stochasticity and thus effect (i) shows how the type resulting in the greatest population size tends to be favoured [14–16]. Likewise, the rate of population turnover (as measured by the neutral variance in per-capita growth rate, $T(p)$) can be thought of as controlling the frequency of steps taken by the random walker. Thus the process will tend to spend more time at values of p that correspond to less frequent steps, and so effect (ii) shows how the type minimizing $T(p)$ tends to be favoured. Taken together these two effects therefore favour the type minimizing the amount of demographic stochasticity, as given by the ratio $T(p)/n(p)$.

We can now examine how the different social traits influence effects (i) and (ii). If the social trait is altruism, then as explained earlier the population size will increase as its frequency increases (i.e., $dn/dp > 0$; this process was the focus of previous work on the role of demographic stochasticity [14–16]). On the other hand, if the trait is spite then the population size will decrease as its frequency increases (i.e., $dn/dp < 0$). Thus effect (i) always favours altruism and disfavours spite. The role played by effect (ii) is more complex. To see why, observe that

on the slow timescale the demographic processes are in quasi-equilibrium and so $T(p) = 2b(p) = 2m(p)$. Therefore if either $b(p)$ or $m(p)$ are constant with respect to p then $dT/dp = 0$. In this case only term (i) plays a role and so altruism is always favoured and spite disfavoured. Otherwise, to understand how the social action affects $T(p)$, we need to consider two cases: (a) the social action affects the death rate, or (b) the social action affects the birth rate.

Consider the case where the social action affects the death rate. If the the social action is altruism then by definition it must decrease the death rate ($dm/dp < 0$) and so we have $dT/dp < 0$. Conversely, if the social action is spite then by definition it must increase the death rate ($dm/dp > 0$) and so $dT/dp > 0$. In both cases effect (ii) works in concert with effect (i) to always favour altruism and disfavour spite. Indeed the ratio $T(p)/n(p)$ is monotonic is p, being minimized at $p = 1$ in the case of altruism and at $p = 0$ in the case of spite.

Next consider the case where the social action affects the birth rate. If the social action is altruism then by definition it must increase the birth rate ($db/dp > 0$) and so we have $dT/dp > 0$. On the other hand, if the social action is spite then by definition it must decrease the birth rate ($db/dp < 0$) and so we have $dT/dp < 0$. Hence effect (ii) opposes effect (i). As a result, altruism or spite can each be favoured or not depending upon the magnitude of effect (i) relative to the magnitude of effect (ii). Moreover, the ratio $T(p)/n(p)$ can be non-monotonic, meaning that it can be minimized by a polymorphic population.

To illustrate these phenomena more concretely, we apply our analysis to several specific models (see S1 Appendix for details). Throughout we use x_i to denote the density of type i.

1. **Social action alters death rate.** Consider a population in which social actors alter the death rate of others. This could be through, for example, the actors producing a diffusible compound such as a resource (e.g., the enzyme invertase in *S. cerevisiae* [16, 19]) or toxin (e.g., bacteriocins [20]). Let $m \equiv d(1 + \nu x_1/[x_1 + x_2])$, with $\nu \in (-1, 1)$. Then the type of social trait is determined by the sign of ν: if $\nu > 0$, the trait is spite, whereas if $\nu < 0$ the trait is altruism. Suppose population size is regulated by density-dependent fecundity, and so let $b \equiv \beta(1 - x_1 - x_2)$, with $\beta > d(1 + |\nu|)$. Thus $T(p)/n(p)$ is decreasing in p if $\nu < 0$ (altruism) and increasing if $\nu > 0$ (spite) (see also S1 Appendix). Fig 1 shows that, as our analysis predicts, demographic stochasticity favours altruism and disfavours spite. It also illustrates how the evolutionary outcome depends upon mutation rate.

2. **Social action alters birth rate.** Here we consider separate models for altruism and spite. For altruism, we suppose that social actors produce a public good that increases growth/reproduction, and that uptake of this good occurs through mass-action contact between the social actor and the recipient. One such example is the production of siderophores by *Pseudomonas aeruginosa* to scavenge iron essential for bacterial growth [35]. As such, we suppose social actors increase the birth rate of others by an amount $\nu > 0$, and so let $b \equiv \beta + \nu x_1$. For spite, suppose individuals attempt to reproduce at a per-capita rate β, and with probability $\nu x_1/(x_1 + x_2 + a)$ reproduction is blocked by a social actor (so $\nu \in [0, 1]$, $a > 0$), and thus $b \equiv \beta(1 - \nu x_1/[x_1 + x_2 + a])$. This could represent a population of sexual hermaphrodites such that when social actors play the role of 'male' they spitefully reduce their investment in gametes, leading to reproductive failure (so a controls probability of self-fertilization). For both models, let the per-capita mortality rate be $m \equiv d + \kappa_1(x_1 + x_2) + \kappa_2(x_1 + x_2)^2$. Fig 2a and 2c shows that, as our analysis predicts, demographic stochasticity can now disfavour altruism and favour spite. In fact, the ratio $T(p)/n(p)$ can be non-monotonic in p and so be minimized by a polymorphic population containing social actors and non-actors. This also suggests that in such cases an intermediate level of social action might, in some sense, be optimal (see S1 Appendix). For example, in a monomorphic population the value of ν minimizing the ratio T/n for the altruism model is $\nu^* = \kappa_1 - |\beta - 2d|/\theta$,

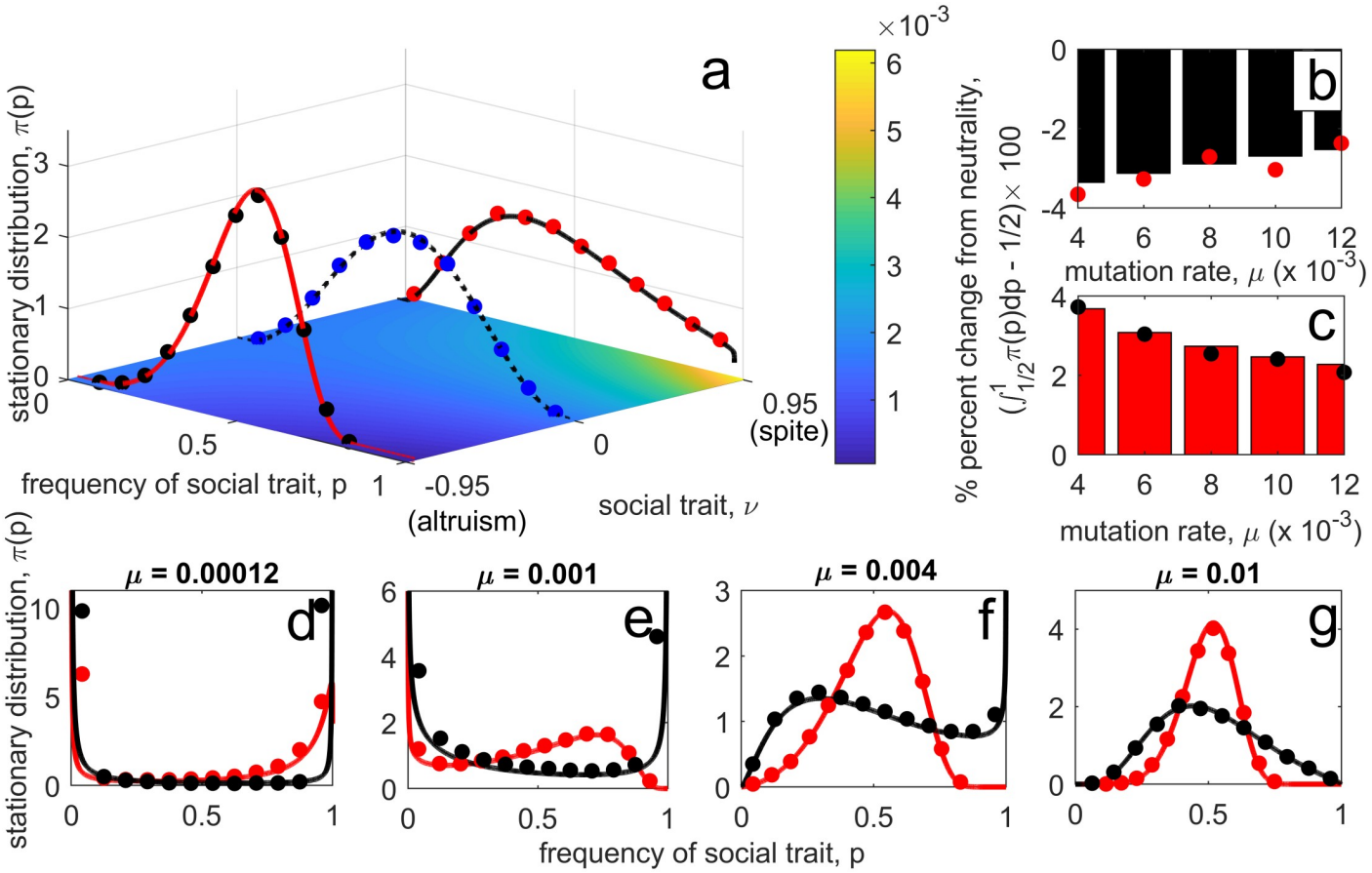

Fig 1. Role of demographic stochasticity in the evolution of cost-free social traits acting on death rate. The model uses $b \equiv 3(1 - x_1 - x_2)$, $m \equiv 1 + \nu x_1/(x_1 + x_2)$ with $\Omega = 900$; if $\nu > 0$, the trait is spite, whereas if $\nu < 0$, the trait is altruism. Subplot **a**—stationary distributions corresponding to three different values of ν: altruism ($\nu = -0.95$), neutral ($\nu = 0$), and spite ($\nu = 0.95$), revealing the close match between our analytic results and simulations of the full stochastic process. Mutation rate is $\mu = 0.006$ in all cases. The distribution is skewed towards a higher frequency of the social actor in the case of altruism and towards a lower frequency of the social actor in the case of spite (distribution is symmetric in the neutral case). Underlying contour plot shows the value of the ratio $T(p)/[\Omega n(p)]$. Subplots **b** and **c** shows the degree to which the social actor is disfavoured (spite, $\nu = 0.95$; subplot **b**) or favoured (altruism, $\nu = -0.95$; subplot **c**) for different mutation rates: as mutation rate decreases, the effect of demographic stochasticity increases. Subplots **d**-**g** show how changing mutation rate alters the shape of the stationary distribution. When mutations are low (subplot **d**), the stationary measure is U-shaped, but skewed in favour of the social actor if the trait is altruism (red curve) or non-social actor if the trait is spite (black curve). As mutation rate increases, the distribution is initially pushed into the interior at the boundary for which the ratio $T(p)/n(p)$ is minimized ($p = 1$ and red curve on subplot **e**; $p = 0$ and black curve on subplot **f**), before the distribution ultimately becomes unimodal with distribution favouring the type which minimizes $T(p)/n(p)$ (subplot **g**). In all plots, the curves/bars are analytic predictions, while circles are the average of 3×10^4 simulations of the full stochastic process (see S1 Appendix). For subplot **b** and **c**, simulations were terminated after 5×10^5 and 7.5×10^5 time units, respectively.

https://doi.org/10.1371/journal.pcbi.1006739.g001

where $\theta = \sqrt{d/\kappa_2}$, and for the spite model is $\nu^* = (a + \theta)(\beta - 2d - \kappa_1\theta)/(\beta\theta)$. Fig 2b and 2d shows that when one type of actor displays this level of social action, all other levels of social action ν are disfavoured.

Simulation results suggest that when more than two types of individuals are included in the population, the above results hold. For example, Fig 3a shows that when the social trait acts on death rate and there are several different types of individuals in the population, ranging from very altruistic to very spiteful, it is the most altruistic type that is favoured. Furthermore, Fig 3b and 3c shows that when the social trait acts on birth rate and there are multiple types of individuals in the population, it can be an intermediate level of altruism or spite that is favoured (analogous to Fig 2b and 2d). Up until this point we have assumed the social trait is cost-free,

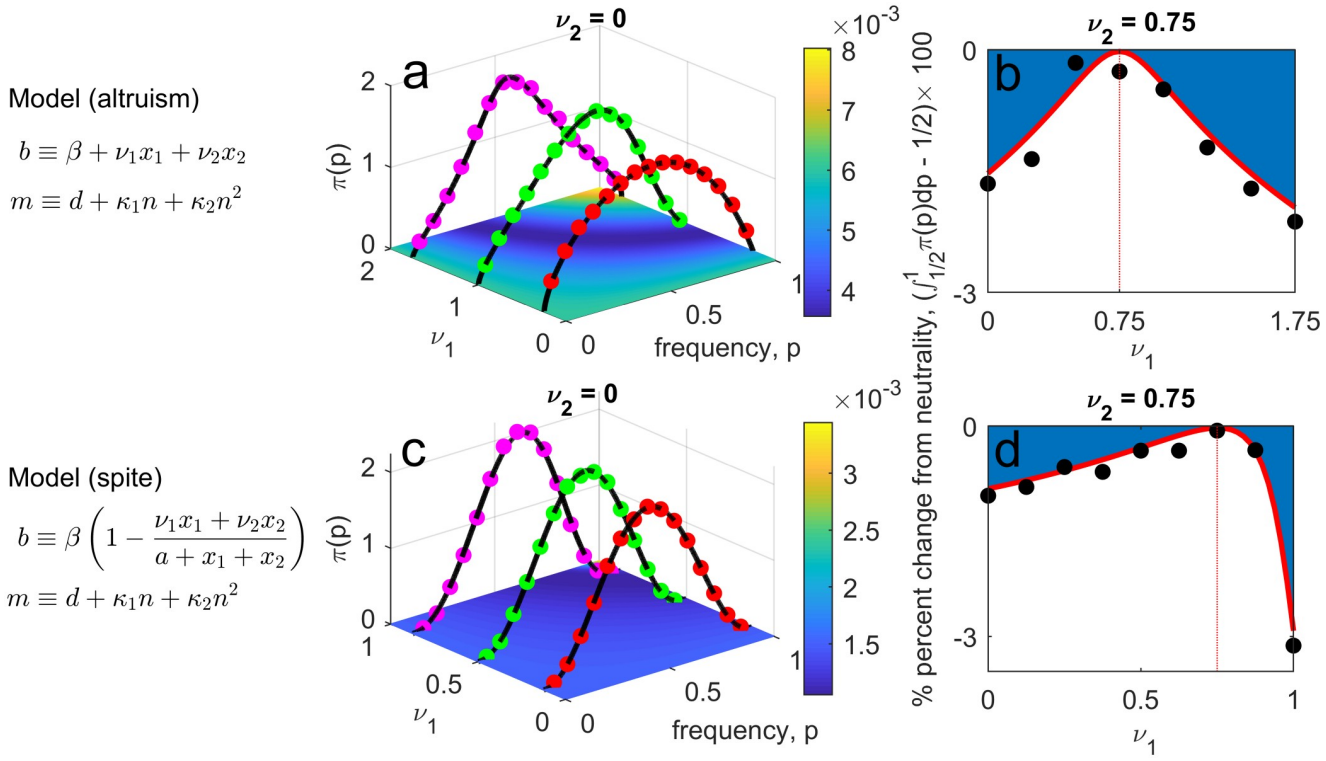

Fig 2. Role of demographic stochasticity in the evolution of cost-free social traits acting on birth rate. In subplots **a**-**b**, the social trait is altruism. In subplots **c**-**d**, the trait is spite. Subplot **a**—stationary distributions corresponding to three different strengths of altruism ν, showing how altruism can be disfavoured. Subplot **c**—stationary distributions corresponding to three different strengths of spite ν, showing how spite can be favoured. Underlying contour plot shows the value of the ratio $T(p)/[\Omega n(p)]$. Subplots **b,d** show that in some cases an intermediate level of social action is optimal. Subplot **b**—a type that uses the intermediate level of altruism that minimizes the ratio $T(p)/n(p)$ in a monomorphic population (in this case $\nu = 0.75$) is favoured over all other levels of altruism. Subplot **d**—a type that uses the intermediate level of spite that minimizes the ratio $T(p)/n(p)$ in a monomorphic population (in this case $\nu = 0.75$) is favoured over all other levels of spite. Curves are analytic predictions and each circle is 6×10^4 simulations of the full stochastic process; simulations were run for 10^3 and 5×10^4 time units for subplots **b** and **d** respectively. Parameters values: $\{\beta, d, \kappa_1, \kappa_2, \Omega, \mu\} = \{1, 0.5, 0.75, 0.01, 250, 0.01\}$ (subplots **a**-**b**) and $\{\beta, d, a, \kappa_1, \kappa_2, \Omega, \mu\} = \{8, 1, 0.05, 0.05, 0.2, 900, 0.005\}$ (subplots **c**-**d**).

https://doi.org/10.1371/journal.pcbi.1006739.g002

$\epsilon = 0$. Suppose instead the social action has a cost, $\epsilon > 0$, which creates a directional bias disfavouring the social trait. We may then ask if/when the effect of stochastic noise can overcome this directional bias, and so reverse the direction of selection [14–16]. We will focus upon situations in which a stationary distribution, $\pi(p)$, exists. Since by construction the social actor (type 1) is at a selective disadvantage ($\epsilon > 0$), if $\int_{1/2}^{1} \pi(p)dp > 1/2$, then we may argue demographic stochasticity reverses the direction of selection.

We illustrate this phenomenon with two examples. First, consider a population where the social actor is an altruist capable of altering birth rate such that $b \equiv r + \nu x_1$, $m \equiv \kappa(x_1 + x_2)$, and $c \equiv r$, where $r > 0$, $\kappa > \nu > 0$. Models based on these specific assumptions have been explored by previous authors, where it was argued that demographic stochasticity favours altruism and thus a selective reversal is possible [14–16]. This argument was based upon two main points. First, the authors observed that the drift term of the SDE associated with the density of social actors, $pn(p)$, could be either positive or negative due to the magnitude of noise-induced effects relative to selection. Second, the authors showed that whichever phenotype can grow to a larger population size in isolation is favoured (altruists) by applying a pairwise comparison of invasion probabilities. Each of these points has an interpretative issue. First, although noise-induced effects often appear in the drift term of the SDE describing the change

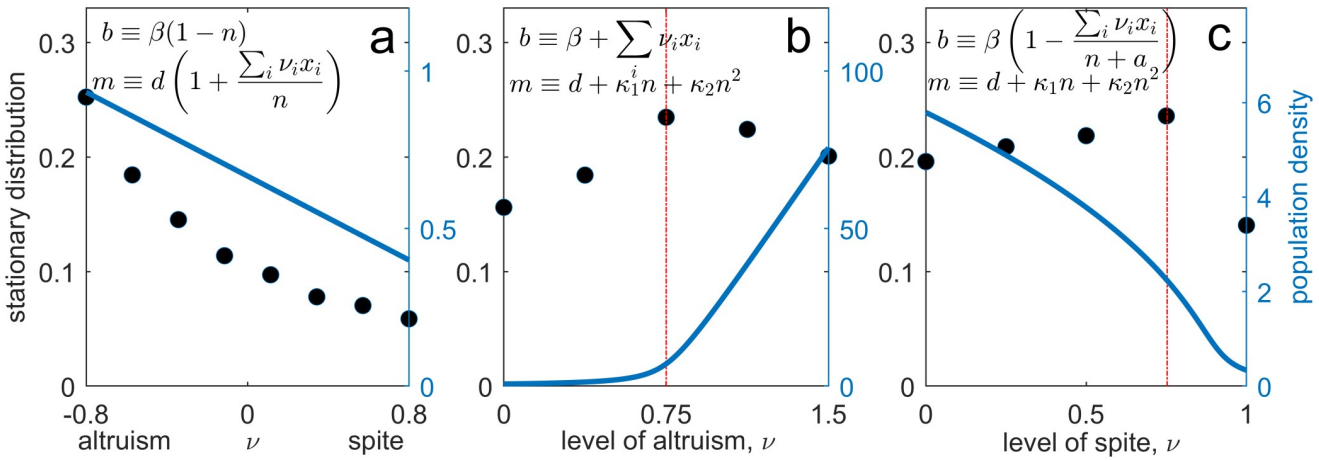

Fig 3. Role of demographic stochasticity in the evolution of cost-free social traits. Each subplot is the model indicated by the per-capita birth and death rates, b and m, with $n = \Sigma_i x_i$. The black circles are the results of 10^4 simulations of the system of SDEs (S1 Appendix) and represent the probability of observing the simulation in a given state (left y-axis). The blue curve is the expected population density of a population monomorphic for the trait value (right y-axis). If population size alone was sufficient to predict which trait is favoured, we would expect a close match between the stationary distribution (black circles) and population size (blue curve)—this does not occur because what is important is the ratio T/n. Indeed, as predicted by consideration of (1), the stochastically favoured trait for subplot **a** is altruism, whereas for subplot **b** and **c** it is the trait value at the red dashed line. Parameter values used: subplot **a**, $\{\beta, d\} = \{3, 1\}$, subplot **b**, $\{\beta, d, \kappa_1, \kappa_2\} = \{1, 0.5, 0.75, 0.01\}$, and subplot **c**, $\{\beta, d, \kappa_1, \kappa_2, a\} = \{8, 1, 0.05, 0.2, 0.05\}$. All simulations used $\Omega = 10^4$ and assumed type i mutates to type j at a per-capita rate $\mu = 10^{-6}$.

https://doi.org/10.1371/journal.pcbi.1006739.g003

in density of social actors (the ecological process), these tend to disappear after the density SDE is converted to the SDE tracking the frequency of the social trait (the evolutionary process), and this is indeed the case here (see Eq 1). It is these noise-induced effects that lead to the incorrect conclusion about when social traits are favoured. To see why, consider the above model when there are no mutations, $\mu = 0$, and no selection, $\epsilon = 0$. Then the social trait (altruism) is neutral. Suppose the population is initially at a state in which half the individuals are social actors, $p = 1/2$. Then since the fixation probability of the social actor in a neutral population is equal to its proportion in the population, 50% of the time the social actor will sweep to fixation in the population. Unsurprisingly, the drift term for the frequency equation, $\alpha(p)$, in Eq 1 is zero, that is, the expected change in frequency is zero. However, the drift term of the SDE for the density of social actors will be positive. This is because the population size goes up when the altruists fix more than it goes down when non-altruists fix. But altruism is neutral, and therefore the sign of the drift term of the density equation cannot be used as a measure of evolutionary 'success'. Second, although comparison of invasion probabilities does favour whichever phenotype grows to a larger population size in isolation, as we pointed out previously, if we place the invasion probabilities within the context of the full mutation-fixation evolutionary process the effect of population size disappears (see also [32–34]).

Indeed, these issues can be made readily apparent by considering the stationary distribution associated with the model (this assumes mutations are explicitly included, which deviates from the model in [16]). In particular, the stationary distribution is

$$\pi(p) \propto p^{\frac{\mu\Omega}{\kappa}-1}(1-p)^{\frac{\mu\Omega}{\kappa}-1}e^{-\frac{\alpha\Omega}{\kappa}p}, \tag{3}$$

(see S1 Appendix). If $\mu\Omega/\kappa > 1$, then mutations push the distribution towards $p = 1/2$ and so $\pi(p)$ has a (skewed) bell-shape, whereas if $\mu\Omega/\kappa < 1$, the distribution accumulates at $p = 0$ and $p = 1$ and so $\pi(p)$ has a (skewed) U-shape. At selective neutrality, $\pi(p)$ is symmetric about $p = 1/2$ and so altruism is completely neutral. If altruism comes at a cost, $\epsilon > 0$, then $\pi(p)$ is

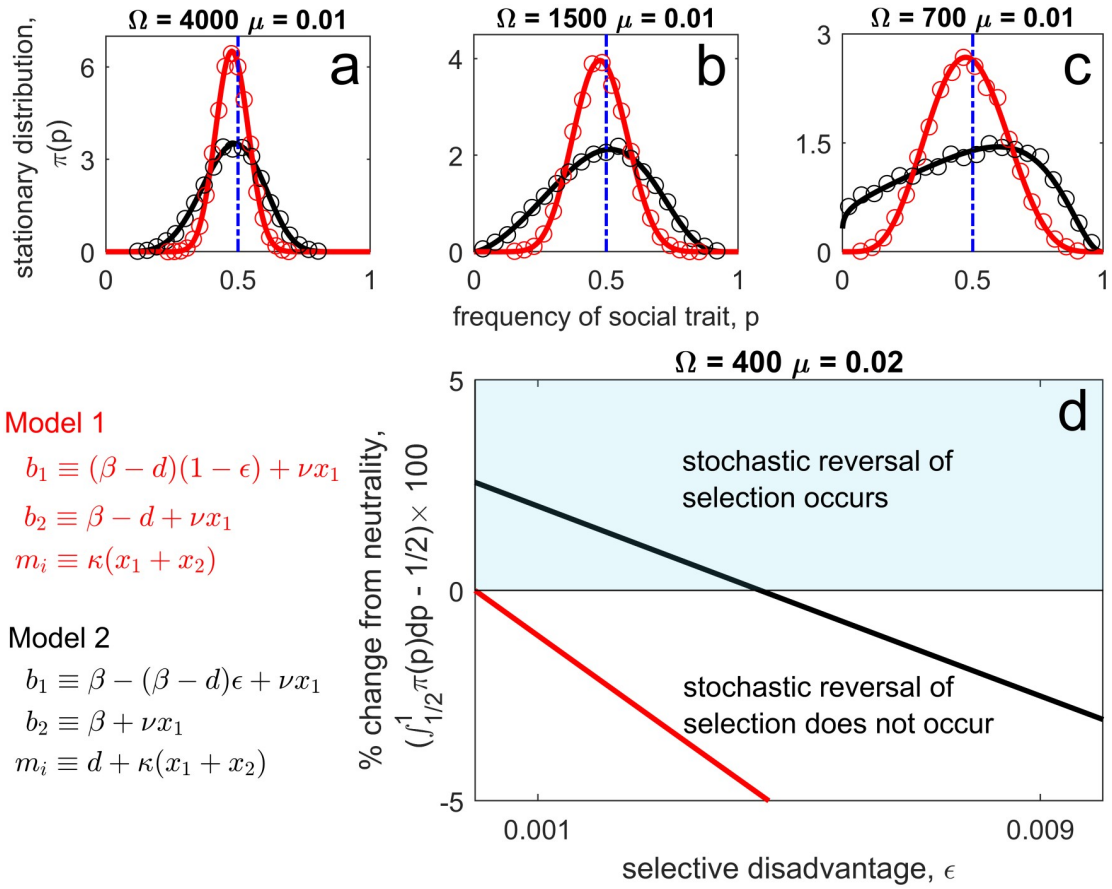

Fig 4. Can stochasticity reverse selection? Here we compare two models of altruism having the same per-capita growth rate, but differing in variance in per-capita growth (Model 1-red, Model 2-black). As a consequence, this can lead to a stochastic reversal of selection for model 2 but not model 1. Subplots **a**-**c**: predicted stationary distribution for both models from Eq 1 (curve) compared to 2×10^4 simulations of the full stochastic process (circles) for decreasing habitat size, **Ω** (i.e., increasing levels of demographic stochasticity). The distribution is always skewed towards the non-altruist in model 1. For model 2 the distribution changes from being skewed towards the non-altruist to being skewed towards the altruist as demographic stochastic increases (i.e., a selective reversal occurs). Subplot **d**: magnitude of the selective reversal plotted against the cost of altruism. Parameters used were $\{\beta, d, \kappa, \nu\} = \{3, 2.4, 1.2, 1\}$, with $\epsilon = 0.003$ for subplots **a**-**c**.

https://doi.org/10.1371/journal.pcbi.1006739.g004

shifted in favour of the non-actor and so stochasticity can never reverse the direction of selection (Fig 4, Model 1—red). This conclusion can also be reached by noting the ratio $T(p)/n(p)$ in this particular model is a constant, independent of p (S1 Appendix). This is because any increase in population size (which reduces the step size of the random walk in p) is exactly compensated for by an increase in the rate of population turnover. Interestingly, it is possible to construct a model in which a selective reversal occurs by making only a slight modification of the above assumptions. Suppose $b \equiv \beta + \nu x_1$, $m \equiv d + \kappa(x_1 + x_2)$, and $c \equiv r$, with $r = \beta - d$. This model has the same per-capita growth rate as the previous model but now the rate of population turnover (i.e., the variance in per-capita growth) is larger. As a result, the ratio $T(p)/n(p)$ is linearly decreasing in p. The stationary distribution is then

$$\pi(p) \propto p^{\frac{\mu\Omega r}{\beta\kappa}-1}(1-p)^{\frac{\mu\Omega r}{\beta\kappa-d\nu}-1}(\beta\kappa - d\nu p)^{\frac{r^2\Omega\epsilon}{d\nu}-\frac{\mu\Omega r}{\beta\kappa}-\frac{\mu\Omega r}{\beta\kappa-d\nu}-1}. \quad (4)$$

In this modified model altruism is now stochastically favoured (Fig 4, Model 2—black) and so stochasticity can reverse the direction of selection. Notice from Eq 4 the role played by

mutation rate in shaping the stationary distribution. In the first model, mutation rate only controlled whether the distribution was normalizable or not. Now, however, mutation rate can alter whether or not a selective reversal is possible.

It is important to stress that the difference in outcome between these two models is driven exclusively by demographic stochasticity. The deterministic components of these two models are the same. Put another way, the expected change in the frequency of the altruists is identical in the two models despite the second model predicting the evolution of costly altruism while the first model not doing so. In the first model selection pushes the distribution in favour of the non-altruists and demographic stochasticity has no biasing effect. In the second model, again selection pushes the distribution in favour of the non-altruists, but now demographic stochasticity is biased such that it decreases as the altruists become more common. The predicted population composition (i.e., the stationary distribution) thus arises from a balance between selection favouring non-altruists and the demographic noise being smaller when the frequency of altruists is high. These effects only become apparent from consideration of the ratio $T(p)/n(p)$. Thus determining whether stochasticity can reverse selection requires analysis of this ratio, and we cannot exclusively focus upon how the social trait alters population size [14–16].

Discussion

Recent work has explored how stochasticity can alter social trait evolution by deriving a stochastic version of Hamilton's rule [17]. Our work differs from this in a couple of important ways. First, those authors focused upon the expected evolutionary change alone, which is equivalent to considering the sign of $\alpha(p)$ of Eq 1, whereas our focus is upon how social traits influence the evolutionary noise, and how this works in conjunction with the expected evolutionary change. Our results demonstrate that examining the expected evolutionary change alone may often be insufficient to determine whether a social trait subject to stochasticity is more or less likely to be observed. Instead one may need to account for both the expected change in the population composition as well as any change in (unbiased) demographic noise that occurs during evolution (i.e., the ratio $T(p)/n(p)$). Second, we have focused on indiscriminate social behaviours and as such, in well-mixed populations these traits are always either neutral (if they are cost-free) or selected against (if they entail a cost). In contrast, Kennedy et al. [17] focuses upon cooperation preferentially directed towards kin.

Our analysis has focused upon unstructured populations in which every individual is equally likely to interact with every other individual. It is well known that population structure can aid or hinder the evolution of social traits [5–7, 36–38] by altering the likelihood that similar or dissimilar social actors interact with one another. Demographic stochasticity will likely factor into this (see in particular [16]), but its impact will depend upon the relatedness of interacting individuals as well as the magnitude of the benefits of the social trait. As relatedness between individuals increases, in general so too will the strength of selection (by generating indirect fitness benefits), which will tend to diminish the role of demographic stochasticity. However, in populations with low relatedness, or social behaviours with sufficiently low benefits (and costs), we would expect our theory to apply. Interestingly, as shown in [16], in deme-structured populations although the social behaviour can be disfavoured or neutral at the within-deme level, it can be favoured at the between-deme level if the social behaviour increases population size and so the number of dispersers [16].

An interesting parallel to our results is that in structured populations, helping behaviours effecting fecundity tend to be selectively favoured over those which effect survivorship [36–39]; a prediction that diverges from our model. One key difference between our model and

these previous studies is that they focused upon the expected change in the social trait in populations of fixed size; as such, whether the helping behaviour is interpreted as one which effects survivorship or one which effects fecundity is based upon whether the population evolves through birth-death or death-birth updating. Hence this result is mediated through the scale of competition between interactants, whereas our result occurs through how the social action effects the evolutionary noise the population experiences.

The role played by demographic stochasticity in populations of fluctuating size has received increased attention recently [14–16, 40, 41]. Our work here has provided a very general consideration of the evolution of two fundamental social traits, altruism and spite, and this analysis has revealed the importance of the action of the social trait upon the recipient. In particular, if the social action alters death rate, then provided selection is sufficiently weak, altruism is stochastically favoured while spite is stochastically disfavoured. If instead the social action alters birth rate, altruism and spite can be either favoured or disfavoured, depending upon mutation rate, the underlying population demography and how this determines the ratio of the rate of population turnover to the population size, $T(p)/n(p)$. The generality of our analysis suggests this principle likely has implications across other study systems as well.

Supporting information

S1 Appendix. Supplementary information. Full derivation of model and details of mathematical analysis.
(PDF)

Author Contributions

Conceptualization: David V. McLeod, Troy Day.

Formal analysis: David V. McLeod, Troy Day.

Funding acquisition: David V. McLeod.

Investigation: David V. McLeod.

Writing – original draft: David V. McLeod.

Writing – review & editing: David V. McLeod, Troy Day.

References

1. Lehmann L, Keller L. The evolution of cooperation and altruism—a general framework and a classification of models. J Evol Biol. 2006; 19:1365–1376. https://doi.org/10.1111/j.1420-9101.2006.01119.x PMID: 16910958
2. Nowak MA. Five rules for the evolution of cooperation. Science. 2006; 314:1560–1563. https://doi.org/10.1126/science.1133755 PMID: 17158317
3. West SA, Gardner A. Altruism, spite, and greenbeards. Science. 2010; 327:1341–1344. https://doi.org/10.1126/science.1178332 PMID: 20223978
4. Frank SA. Foundations of Social Evolution. Princeton, N.J.: Princeton University Press; 1998.
5. Nowak MA, May RM. Evolutionary games and spatial chaos. Nature. 1992; 359:826–829. https://doi.org/10.1038/359826a0
6. Taylor PD. Altruism in viscous populations—an inclusive fitness model. Evol Ecol. 1992; 6:352–356. https://doi.org/10.1007/BF02270971
7. Hauert C, Doebeli M. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. Nature. 2004; 428:643–646. https://doi.org/10.1038/nature02360 PMID: 15074318
8. Allen B, Lippner G, Chen Y, Fotouhi B, Momeni N, Yau S, et al. Evolutionary dynamics on any population structure. Nature. 2017; 544:227–230. https://doi.org/10.1038/nature21723 PMID: 28355181

9. Hamilton WD. The genetical evolution of social behaviour, I and II. J Theor Biol. 1964; 7:1–52. https://doi.org/10.1016/0022-5193(64)90038-4 PMID: 5875341
10. West SA, Griffin AS, Gardner A. Evolutionary explanations for cooperation. Current Biology. 2007; 17: R661–R672. https://doi.org/10.1016/j.cub.2007.06.004 PMID: 17714660
11. Griffin AS, West SA. Kin discrimination and the benefit of helping in cooperatively breeding vertebrates. Science. 2003; 302:634–636. https://doi.org/10.1126/science.1089402 PMID: 14576431
12. Antal T, Ohtsuki H, Wakeley J, Taylor PD, Nowak MA. Evolution of cooperation by phenotypic similarity. Proc Natl Acad Sci. 2009; 106:8597–8600. https://doi.org/10.1073/pnas.0902528106 PMID: 19416902
13. Gardner A, West SA. Greenbeards. Evolution. 2010; 64:25–38. https://doi.org/10.1111/j.1558-5646.2009.00842.x PMID: 19780812
14. Houchmandzadeh B, Vallade M. Selection for altruism through random drift in variable size populations. BMC Evol Biol. 2012; 12. https://doi.org/10.1186/1471-2148-12-61 PMID: 22574999
15. Houchmandzadeh B. Fluctuation driven fixation of cooperative behavior. Biosystems. 2015; 127:60–66. https://doi.org/10.1016/j.biosystems.2014.11.006 PMID: 25451769
16. Constable GWA, Rogers T, McKane AJ, Tarnita CE. Demographic noise can reverse the direction of deterministic selection. Proc Nat Acad Sci. 2016;. https://doi.org/10.1073/pnas.1603693113
17. Kennedy P, Higginson AD, Radford AN, Summer S. Altruism in a volatile world. Nature. 2018; 555:359–362. https://doi.org/10.1038/nature25965 PMID: 29513655
18. Doebeli M, Ispolatov Y, Simon B. Towards a mechanistic foundation of evolutionary theory. Elife. 2017; 15.
19. MacLean RC, Brandon C. Stable public goods cooperation and dynamic social interactions in yeast. J Evol Biol. 2008; 21:1836–1843. https://doi.org/10.1111/j.1420-9101.2008.01579.x
20. Riley MA, Wertz JE. Bacteriocins: evolution, ecology, and application. Annu Rev Microbiol. 2002; 56:117–137. https://doi.org/10.1146/annurev.micro.56.012302.161024 PMID: 12142491
21. West SA, Griffin AS, Gardner A, Diggle SP. Social evolution theory for microorganisms. Nat Rev Microbiol. 2006; 4:597–607. https://doi.org/10.1038/nrmicro1461 PMID: 16845430
22. West SA, Griffin AS, Gardner A. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J Evol Biol. 2007; 20:415–432. https://doi.org/10.1111/j.1420-9101.2006.01258.x PMID: 17305808
23. Gardiner CW. Handbook of Stochastic Methods. Berlin, Germany: Springer: Complexity; 2009.
24. Allen LJS. An Introduction to Stochastic Processes with Applications to Biology. Boca Raton, FL: CRC Press; 2011.
25. Parsons TL, Rogers T. Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: a constructive approach with examples from theoretical biology. J Phys A: Mathematical and Theoretical. 2017; 50. https://doi.org/10.1088/1751-8121/aa86c7
26. Berglund N, Gentz B. Noise-induced Phenomena in Slow-Fast Dynamical Systems. USA: Springer; 2006.
27. Constable GWA, McKane AJ, Rogers T. Stochastic dynamics on slow manifolds. J Phys A: Math Theor. 2013; 46. https://doi.org/10.1088/1751-8113/46/29/295002
28. Katzenberger GS. Solutions of a stochastic differential equation forced onto a manifold by a large drift. Ann Probab. 1991; 19:1587–1628. https://doi.org/10.1214/aop/1176990225
29. Karlin S, Taylor HM. A Second Course in Stochastic Processes. New York, NY: Academic Press; 1981.
30. Ewens WJ. Mathematical Population Genetics I: Theoretical Introduction. New York, NY: Springer: Interdisciplinary Applied Mathematics; 2004.
31. Gillespie JH. The Causes of Molecular Evolution. Oxford, UK: Oxford University Press; 1991.
32. McCandlish DM, Stoltzfus A. Modeling evolution using the probability of fixation: history and implications. Quarterly Review of Biology. 2014; 89:225–252. https://doi.org/10.1086/677571 PMID: 25195318
33. Kimura M. Evolutionary rate at the molecular level. Nature. 1968; 217:624–626. https://doi.org/10.1038/217624a0 PMID: 5637732
34. King JL, Jukes TH. Non-Darwinian evolution. Science. 1969; 164:788–798. https://doi.org/10.1126/science.164.3881.788 PMID: 5767777
35. Griffin AS, West SA, Buckling A. Cooperation and competition in pathogenic bacteria. Nature. 2004; 430:1024–1027. https://doi.org/10.1038/nature02744 PMID: 15329720

36. Ohtsuki H, Hauert C, Lieberman E, Nowak MA. A simple rule for the evolution of cooperation on graphs and social networks. Nature. 2006; 441:502–505. https://doi.org/10.1038/nature04605 PMID: 16724065

37. Taylor PD, Day T, Wild G. Evolution of cooperation in a finite homogeneous graph. Nature. 2007; 447:469–472. https://doi.org/10.1038/nature05784 PMID: 17522682

38. Debarre F, Hauert C, Doebeli M. Social evolution in structured populations. Nat Comm. 2014; 5. https://doi.org/10.1038/ncomms4409

39. Lehmann L, Rousset F. How life history and demography promote or inhibit the evolution of helping behaviours. Phil Trans Roy Soc B. 2010; 365:2599–2617. https://doi.org/10.1098/rstb.2010.0138

40. Hallatschek O. Noise driven evolutionary waves. PLOS Comp Biol. 2011; 7. https://doi.org/10.1371/journal.pcbi.1002005

41. Hallatschek O. The noisy edge of travelling waves. Proc Natl Acad Sci. 2011; 108:1783–1787. https://doi.org/10.1073/pnas.1013529108 PMID: 21187435

Supporting information for "Social evolution under demographic stochasticity"

David V. McLeod[1*], Troy Day[2†]

1 Institute for Integrative Biology, ETH Zürich, Zürich, Switzerland
2 Department of Mathematics and Statistics, Department of Biology Queen's University, Kingston, ON, Canada

* david.mcleod@env.ethz.ch
† day@queensu.ca

General model

We first derive the general population model by specifying a continuous time discrete state space Markov process. To do so, consider a population consisting of two types of individuals and potentially some environmental variable that the two types interact with (e.g., a diffusible compound produced by type 1). Let X_i denote the number of individuals of type $i = 1, 2$ at time t and Y denote the state of the environmental variable. Let $\mathbf{X} = (X_1, X_2, Y)$. The infinitesimal transition probabilities regulating the population demographics are

Event type	Transition	Probability	
birth of type i	$\mathbf{X} \to \mathbf{X} + \mathbf{e}_i$	$\mathbb{T}_i(\mathbf{X}) \equiv b_i(\mathbf{X})X_i\Delta t,$	$i, j = 1, 2,$
death of type i	$\mathbf{X} \to \mathbf{X} - \mathbf{e}_i$	$\mathbb{T}_{i+2}(\mathbf{X}) \equiv m_i(\mathbf{X})X_i\Delta t,$	$i \neq j$
mutation of type i to j	$\mathbf{X} \to \mathbf{X} - \mathbf{e}_i + \mathbf{e}_j$	$\mathbb{T}_{i+4}(\mathbf{X}) \equiv \mu X_i \Delta t$	

(S1)

where $\mathbf{e}_i$ is the 1×3 vector with a 1 in the i-th spot and zeros in the others. We do not explicitly specify the dynamics of the environmental variable for reasons which will become apparent shortly, but suppose that it may under go any possible transition of the form $\mathbf{X} \to \mathbf{X} \pm \mathbf{e}_3$ or $\mathbf{X} \to \mathbf{X} \pm \mathbf{e}_3 \mp \mathbf{e}_i$. In the latter case, the effect upon type 1 or 2 is subsumed into the birth and death terms of eq S1. Let $Q(\mathbf{X}, t)$ be the probability density for $\mathbf{X}$ at time t. Then the master equation for the stochastic process (ignoring changes in Y) is

$$\frac{\partial Q}{\partial t} = \sum_{i=1}^{2} \left[\mathbb{T}_i(\mathbf{X} - \mathbf{e}_i)Q(\mathbf{X} - \mathbf{e}_i, t) + \mathbb{T}_{i+2}(\mathbf{X} + \mathbf{e}_i)Q(\mathbf{X} + \mathbf{e}_i, t) \right] + \sum_{\substack{i,j=1,2 \\ i \neq j}} \mathbb{T}_{i+4}(\mathbf{X} + \mathbf{e}_i - \mathbf{e}_j)Q(\mathbf{X} + \mathbf{e}_i - \mathbf{e}_j, t) - \sum_{k=1}^{6} \mathbb{T}_k(\mathbf{X})Q(\mathbf{X}, t). \quad \text{(S2)}$$

Let Ω be a system size parameter (e.g., [1, 2]), and take $\mathbf{x} \equiv (x_1, x_2, y) \equiv \mathbf{X}/\Omega$, and $q(\mathbf{x}, t) \equiv \Omega Q(\mathbf{X}, t)$. In our context Ω can be thought of as habitat size and so $\mathbf{x}$ can be thought of as densities. If we suppose Ω is sufficiently large such that the variables $\mathbf{x}$ are approximately continuous and rescale time as $\tau = t/\Omega$, we can perform a series expansion in powers of $1/\Omega$ to give the Fokker-Planck (or forward Kolmogorov) equation

$$\frac{\partial q}{\partial \tau} = -\sum_{i=1}^{2} \frac{\partial}{\partial x_i} A_i(\mathbf{x})q(\mathbf{x}, \tau) + \frac{1}{2\Omega} \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\partial^2}{\partial x_i x_j} B_{ij}(\mathbf{x})q(\mathbf{x}, \tau) \quad \text{(S3)}$$

where $A_i(\mathbf{x}) = (b_i(\mathbf{x}) - m_i(\mathbf{x}))x_i - \mu(x_i - x_j)$, and the $B_{ij}(\mathbf{x})$ are the entries of

$$\boldsymbol{B}(\mathbf{x}) = \begin{pmatrix} (b_1(\mathbf{x}) + m_1(\mathbf{x}))x_1 + \mu n & -\mu n \\ -\mu n & (b_2(\mathbf{x}) + m_2(\mathbf{x}))x_2 + \mu n \end{pmatrix}, \quad \text{(S4)}$$

with $n = x_1 + x_2$.

The Fokker-Planck equation given by eq S3 is associated with the system of Ito stochastic differential equations (SDEs)

$$d\mathbf{x} = \boldsymbol{A}(\mathbf{x})d\tau + \Omega^{-1/2}\boldsymbol{C}(\mathbf{x})d\mathbf{W}_\tau \quad \text{(S5)}$$

where $\boldsymbol{A}(\mathbf{x}) = (A_1(\mathbf{x}), A_2(\mathbf{x}))^T$, $\boldsymbol{C}(\mathbf{x})\boldsymbol{C}(\mathbf{x})^T = \boldsymbol{B}(\mathbf{x})$ and $\mathbf{W}_\tau$ is a vector of N independent Wiener processes [1, 3]. Thus $\boldsymbol{C}(\mathbf{x})$ is a $2 \times N$ matrix, where N is a integer whose value will depend upon how the matrix $\boldsymbol{C}(\mathbf{x})$ is chosen (the choice of matrix $\boldsymbol{C}(\mathbf{x})$ is not

unique). If habitat size becomes large, $\Omega \to \infty$, stochasticity disappears from eq S5 and we are left with the system of ordinary differential equations (ODEs) $\dot{\mathbf{x}} = \boldsymbol{A}(\mathbf{x})$, where $\dot{\mathbf{x}} = (dx_1/d\tau, dx_2/d\tau)^T$ (we have purposefully neglected $dy/d\tau$ in $\dot{\mathbf{x}}$).

As in the main text, we assume that type 1 is the social actor, and so if we let $r_i(\mathbf{x}) \equiv b_i(\mathbf{x}) - m_i(\mathbf{x})$ denote the per-capita growth rate of type i, then we assume that $r_1(\mathbf{x}) = b(\mathbf{x}) - m(\mathbf{x}) - \epsilon c(\mathbf{x})$ and $r_2(\mathbf{x}) = b(\mathbf{x}) - m(\mathbf{x})$. Thus whenever $\epsilon > 0$, type 2 has the selective advantage, and will ultimately fix in the ODE model without mutations. Note that the costs may either reduce birth rate or increase death rate; although in general this will have implications for the structure of $\boldsymbol{B}(\mathbf{x})$ and thus $\boldsymbol{C}(\mathbf{x})$, because we will assume ϵ is sufficiently small that we can neglect terms of order ϵ/Ω, we only need to take into account how the costs alter expected per-capita growth rate.

Reduction of system to slow manifold

We wish to reduce system S5 into a more manageable problem. To do so, suppose that in the absence of selection, mutations, and stochasticity ($\epsilon = 0$, $\mu = 0$, and $\Omega \to \infty$, respectively), there exists a globally asymptotically stable curve of ecological equilibria in the ODE system given by $r_i(\mathbf{x}) = 0$. We parameterize this curve in terms of x_1, and so let $\boldsymbol{\gamma}(w) = (w, \gamma_2(w), \gamma_y(w))$ denote the value of $\mathbf{x}$ along the curve. We will assume that both $w/\gamma_2(w)$ and $w/(w + \gamma_2(w))$ are invertible. When $\mu = \epsilon = 0$ and $\Omega \to \infty$, the ODE system will asymptotically approach different points on $\boldsymbol{\gamma}(w)$ dependent upon the initial conditions ($\boldsymbol{\gamma}(w)$ is a center manifold [4]). If instead Ω is large but finite, and ϵ, μ are nonzero but small, then system S5 will rapidly approach $\boldsymbol{\gamma}(w)$ along the flow lines of the ODE system. However, once in the vicinity of $\boldsymbol{\gamma}(w)$ the stochastic component of system S5 will dominate the dynamics, since as $\mathbf{x} \to \boldsymbol{\gamma}(w)$, $\boldsymbol{A}(\mathbf{x}) \to 0$. Because the movement along $\boldsymbol{\gamma}(w)$ is slow relative to the rate at which a system initially distant moves to the vicinity of $\boldsymbol{\gamma}(w)$, in this context $\boldsymbol{\gamma}(w)$ is often referred to as a 'slow manifold' [5–7]. Since the movement along the slow manifold corresponds to change in population composition (and so represents the evolutionary timescale), our goal is to derive an equation approximating the motion of system S5 along $\boldsymbol{\gamma}(w)$.

To do so, observe that when $\epsilon = \mu = 0$ and $\Omega \to \infty$ we have

$$\frac{dx_1/d\tau}{dx_2/d\tau} = \frac{A_1(\mathbf{x})}{A_2(\mathbf{x})} = \frac{x_1}{x_2} \quad \Rightarrow \quad \frac{dx_1}{dx_2} = \frac{x_1}{x_2}, \tag{S6}$$

and so for the initial condition $\mathbf{x}_w = (w_1, w_2, w_3)$, the solution of eq S6 is $x_1(t) = (w_1/w_2)x_2(t)$. Now suppose (stochastic) trajectories along the slow manifold receive random "kicks" displacing the system from the slow manifold. Once the system is "kicked away" from the slow manifold, then provided Ω is large (and ϵ, μ small), the system will return to the slow manifold approximately along the (deterministic) flow lines. These flow lines are the solutions of eq S6 with the initial condition being the position system was "kicked" too. Thus a trajectory initially at (x_1, x_2, y) will return to the point on the slow manifold, $\boldsymbol{\gamma}(w)$, implicitly given by

$$\gamma_2(w) = \frac{x_2}{x_1} w. \tag{S7}$$

From eq S7, since w parameterizes the slow manifold, dw/dt reveals how the position of the system along the slow manifold evolves with time [6, 8]. Let $\mathcal{G}(w) = w/\gamma(w)$ and $g(w) = \mathcal{G}^{-1}(w)$. Then from eq S7 we have $w = g(x_1/x_2)$ and so applying the

multivariable version of Ito's formula [1, 3] to w gives

$$dw = \left[\sum_{i=1}^{2} \frac{\partial g}{\partial x_i} A_i(\mathbf{x}) + \frac{b(\mathbf{x}) + m(\mathbf{x})}{2\Omega} \sum_{i=1}^{2} \frac{\partial^2 g}{\partial x_i^2} x_i\right] d\tau + \sqrt{\sum_{i=1}^{2} \left(\frac{\partial g}{\partial x_i}\right)^2 \frac{(b(\mathbf{x}) + m(\mathbf{x}))x_i}{\Omega}} dW_\tau \tag{S8}$$

where we have neglected terms of order ϵ/Ω and μ/Ω.

Let $p \equiv w/(w + \gamma_2(w))$ be the frequency of type 1 along the slow manifold. Then applying Ito's formula to compute dp using eq S8 and evaluating the result on the slow manifold (see [6, 8] for justification) gives

$$dp = [\mu(1-2p) - \epsilon c(p)p(1-p)]d\tau + \sqrt{\frac{p(1-p)T(p)}{\Omega n(p)}} dW_\tau \tag{S9}$$

with $T(p) = b(\boldsymbol{\gamma}(h(p))) + m(\boldsymbol{\gamma}(h(p)))$, $c(p) = c(\boldsymbol{\gamma}(h(p)))$, $n(p) = h(p) + \gamma_2(h(p))$, where $h(p)$ is defined as $h^{-1}(w) = w/(w + \gamma_2(w))$.

To obtain eq S9 we did not need to specify the dynamics of the environmental variable y, and instead only needed to know the density of the environmental variable along the slow manifold, $\gamma_y(w)$. This is because the separation of time scales assumes population composition changes on a slower time scale than demographic processes and so on the slow timescale the environmental variable will be in a quasi-steady state if we use the tools of Parsons and Rogers [6].

As an aside, we note that for problems which satisfy condition eq S6, then eq S9 is the same result we would have obtained if we instead applied Ito's formula to $p = x_1/(x_1+x_2)$ and then restricted the resulting equation to $\boldsymbol{\gamma}(w)$. The intuitive reason for why this is true is that a system (stochastically) perturbed away from the slow manifold will return roughly along the flow lines of the ODE system when $\epsilon = 0$. But along these flow lines, by eq S6 we see that the proportion p remains constant.

To summarize our analysis to this point, we first specified a discrete state space stochastic process involving two types of competing individuals in eq S1. We then assumed that habitat size (and so population size) was sufficiently large that the discrete state space is approximately continuous, and obtained a diffusion approximation of the stochastic process (eq S3). We then applied the techniques of Parsons and Rogers [6] to derive a single-variable SDE approximating the dynamics of motion of eq S3 along the slow manifold by eliminating the dynamics on the fast timescale.

Speed measure and stationary distribution

As in the main text, let $\alpha(p) \equiv \mu(1-2p) - \epsilon c(p)p(1-p)$ and $\sigma^2(p) \equiv p(1-p)T(p)/[\Omega n(p)]$. Then $\alpha(p)$ and $\sigma^2(p)$ are the infinitesimal mean and variance, respectively, of a one-dimensional diffusion process [9, 10]; this is the diffusion process associated with the SDE S9. Define

$$\pi(p) \equiv \frac{C_0}{\sigma^2(p)} \exp\left(2\int^p \frac{\alpha(z)}{\sigma^2(z)} dz\right) \tag{S10}$$

for some constant C_0. If C_0 can be chosen such that $\int_0^1 \pi(p)dp = 1$, then $\pi(p)$ is normalizable and so represents the stationary distribution of the diffusion process with infinitesimal mean $\alpha(p)$ and infinitesimal variance $\sigma^2(p)$ [1, 3, 10]. However, regardless of whether such a C_0 exists, $\pi(p)$ is the speed measure of the diffusion process [9]. The relevance of this is that the speed measure is proportional to the expected time a diffusion process initially at p takes to exit an interval $(\varepsilon - p, p + \varepsilon)$ for small ε [9]. Thus as $\pi(p)$

increases, the process will tend to spend more time at state p, and so we are more likely to observe the process in such a state. Hence as $\pi(p)$ increases for a given p, we will say state p is increasingly favoured.

There are three factors that influence eq S10

1. selection, which is controlled by $\epsilon c(p)$, and biases the stochastic process against the social actor,

2. mutation rate, μ, which pushes $\pi(p)$ away from the boundaries ($p = 0$ and $p = 1$) towards $p = 1/2$,

3. and demographic stochasticity, which is controlled by the infinitesimal variance, $\sigma^2(p)$. If we inspect $\sigma^2(p)$, however, we see that it is the product of a symmetric term, $p(1-p)/\Omega$, and the ratio, $T(p)/n(p)$. Thus this ratio controls the effect of demographic stochasticity, and as we will see is key to understanding social trait evolution.

Evolution of cost-free social traits

To understand the role of demographic stochasticity, suppose that selection is turned off ($\epsilon = 0$). Then the only two factors present in eq S10 are mutation rate and the ratio $T(p)/n(p)$. First, consider how $T(p)/n(p)$ changes in p, that is,

$$\frac{d}{dp}\left[\frac{T(p)}{n(p)}\right] = \frac{T(p)}{n(p)}\left[\frac{dT/dp}{T(p)} - \frac{dn/dp}{n(p)}\right]. \tag{S11}$$

If the social trait is spite, $dn/dp < 0$, whereas if the social trait is altruism, $dn/dp > 0$. Since on the slow manifold, $T(p) = 2b(p) = 2m(p)$, if the social trait acts upon death rate, then $dT/dp = 2dm/dp > 0$ if the trait is spite, whereas if the trait is altruism, $dT/dp < 0$. Thus for social traits acting on the death rate, $T(p)/n(p)$ is monotonic in p, and so is either minimized by the social actor (if the trait is altruism) or the non-social actor (if the trait is spite). If instead the social trait acts upon birth rate, then since $dT/dp = 2db/dp$, if the trait is spite $dT/dp < 0$ and if the trait is altruism $dT/dp > 0$. Hence for both altruism and spite the ratio $T(p)/n(p)$ may be smaller for a population monomorphic for the social actor, or it may be smaller for a population monomorphic for the non-social actor. Moreover, it is also possible that $T(p)/n(p)$ is non-monotonic in p (and so minimized by a polymorphic population).

To understand the implications of the behaviour of $T(p)/n(p)$, first we will assume that $T(p)/n(p)$ is monotonic in p, and so is minimized by either type 1 or type 2, before considering what happens when $T(p)/n(p)$ is non-monotonic in p. There are three different regimes based upon mutation rate, and we consider each in turn.

1. **Low mutation rate**. When mutation rate is sufficiently low, then in the interval $p \in (0, 1)$, $\pi(p) \approx 1/\sigma^2(p)$, and $\pi(p)$ is U-shaped. We are interested in whether $\pi(p)$ is increasing or decreasing in p. Taking the derivative gives

$$\frac{d\pi}{dp} \approx \frac{d}{dp}\left[\frac{1}{\sigma^2(p)}\right] = -\frac{1}{\sigma^2(p)}\left(\frac{(1-2p)}{p(1-p)} + \frac{n(p)}{T(p)}\frac{d}{dp}\left[\frac{T(p)}{n(p)}\right]\right). \tag{S12}$$

   Since the first term, $(1-2p)/[p(1-p)]$, is symmetric in p, we can ignore it and instead focus upon how the ratio $T(p)/n(p)$ changes with respect to p. If the ratio is decreasing (resp. increasing) in p, then $d\pi/dp > 0$ (resp. $d\pi/dp < 0$), and the social actor is favoured (resp. disfavoured).

   An alternative way to arrive at this conclusion is to note that in the absence of selection (and mutations), the diffusion process is on its natural scale (i.e.,

$\alpha(p) = 0$), and so the fixation probability of either type is simply equal to its proportion in the population. Suppose there are k possible types of individuals, and let mutations be sufficiently rare such that between mutations the population returns to a monomorphic state. Then we can construct a Markov chain on the state space of possible strain types [11]. In particular, let μ_{ij} be the rate at which strain i mutates to strain j and N_i be the number of type i individuals in the population at the moment of the (rare) mutation. Then

$$M(i,j) = (\text{mutation rate}) \times (\text{fixation probability}) = \mu_{ij} N_i \times \frac{1}{N_i} = \mu_{ij} \qquad \text{(S13)}$$

is the rate at which the population transitions from a monomorphic strain i state to a monomorphic strain j state [11]. Hence in the absence of mutational biases, the Markov chain is equally likely to be in any particular state [11–13]. This implies that the process will spend equal time in any of the monomorphic states. But if so, then what type we are most likely to observe in the population will be dictated by the time the process takes to transition between states, that is, the expected time from the initial appearance of a mutation for the population to return to a monomorphic state. Using standard techniques [10], when there are two types in the population and they are selectively neutral, $\epsilon = 0$, then the expected time till absorption (time to reach either $p = 0$ or $p = 1$) from an initial state p_0 is

$$\bar{t}(p_0) = 2(1-p_0)\int_0^{p_0} \frac{\Omega n(p)}{(1-p)T(p)} dp + 2p_0 \int_{p_0}^1 \frac{\Omega n(p)}{pT(p)} dp. \qquad \text{(S14)}$$

Since time spent in any monomorphic state is equal, the type that is most likely to be observed is the type which maximizes absorption time, that is, the type that maximizes the time spent transitioning between states.

To determine this, let ν control the effect of the social behaviour, and suppose that the effect of the social behaviour upon the ratio $T(p)/n(p)$ is small. Since the two types only differ due to the effects controlled by ν, we can write $T(p)/n(p) = T(\nu p)/n(\nu p)$, and so a Taylor expansion of the ratio $T(p)/n(p)$ with respect to ν gives

$$\frac{T(p)}{n(p)} = \frac{T}{n} + \nu p \left[\frac{\partial}{\partial[\nu p]} \frac{T(\nu p)}{n(\nu p)} \right]_{\nu=0} + \mathcal{O}(\nu^2). \qquad \text{(S15)}$$

If we then use this approximation for the ratio $T(p)/n(p)$ in the differential equation used to derive eq S14, we obtain the time till absorption as

$$\bar{t}(p_0) = -\frac{2\Omega}{T/n} \left(\ln([1-p_0]^{1-p_0} p_0^{p_0}) - \frac{\nu}{T/n} \frac{\partial T/n}{\partial[\nu p]} \ln([1-p_0]^{1-p_0}) \right) + \mathcal{O}(\nu^2), \qquad \text{(S16)}$$

where T/n no longer depends upon ν or p (or p_0). Notice from eq S16 that if the two types are identical in every respect, $\nu = 0$, then absorption time is symmetric in p about $p = 1/2$, and for any p, $\bar{t}(p)$ will decrease (resp. increase) as T/n becomes large (resp. small). Thus absorption time is maximized (resp. minimized) when we have minimized T/n (resp. maximized T/n). If instead $\nu > 0$, then time till absorption will no longer be symmetric about $p = 1/2$, and instead will depend upon how the social behaviour effects the ratio T/n. In particular, if we consider the difference $\bar{t}(p_0) - \bar{t}(1-p_0)$ on the interval $p_0 \in (0, 1/2)$, we see that this quantity will be negative, that is, the process will take longer to absorb from state $1 - p_0$ then p_0, provided the social trait minimizes the ratio T/n. In this

circumstance, it follows immediately that $\int_{1/2}^{1} \bar{t}(p_0)dp_0 / \int_{0}^{1} \bar{t}(p_0)dp_0 > 1/2$. Hence if the social trait minimizes the ratio T/n, then absorption time will be biased in favour of the social actor and so the type minimizing T/n is favoured when mutations are sufficiently rare.

2. **Intermediate mutation rate**. Suppose mutation rate is non-negligible, but is not sufficiently high so as to admit a normalizable stationary distribution. Now the picture is more complex, and it is not clear how to determine which type is favoured. To see why, consider how $\pi(p)$ changes in p:

$$\begin{aligned}\frac{d\pi}{dp} &= 2\frac{\pi(p)}{\sigma^2(p)}\left(\alpha(p) - \frac{1}{2}\frac{d\sigma^2(p)}{dp}\right), \\ &= 2\frac{\pi(p)}{\sigma^2(p)}\Bigg(\underbrace{(1-2p)\left(\mu - \frac{T(p)}{2\Omega n(p)}\right)}_{(1)} - \underbrace{\frac{p(1-p)}{2\Omega}\frac{d}{dp}\left[\frac{T(p)}{n(p)}\right]}_{(2)}\Bigg).\end{aligned} \tag{S17}$$

From how the terms are grouped, we see that as $p \to 1/2$ term (1) disappears, whereas when $p \to 0$ or $p \to 1$, term (2) disappears (note it is also possible that there exists a p_0 such that $\mu = T(p_0)/[2\Omega n(p_0)]$, in which case term (1) also disappears as $p \to p_0$). Roughly speaking, this implies that term (1) is the stronger effect near the boundaries of the interval $p \in [0, 1]$, whereas term (2) is the stronger effect in the interior. Term (1) represents the effect of mutations, μ, which pushes $\pi(p)$ towards $p = 1/2$, and genetic drift, $T(p)/[2\Omega n(p)]$, which pushes $\pi(p)$ towards the boundaries. These effects would be present even if the ratio $T(p)/n(p)$ were constant, that is, type 1 and 2 were mathematically interchangeable. Term (2), however, occurs due to how the magnitude of demographic stochasticity (or genetic drift) changes due to the consequences of the social trait, that is, how the ratio $T(p)/n(p)$ changes in p.

Consider the behaviour of $\pi(p)$ at the boundaries, and so focus upon term (1). When mutation rate satisfies $\mu < \frac{T(p)}{2\Omega n(p)}$ for all p, then the distribution tends to accumulate at both boundaries since genetic drift is stronger then mutations. As μ increases, however, because in general $\frac{T(0)}{2\Omega n(0)} \neq \frac{T(1)}{2\Omega n(1)}$, there will be a regime in which $\mu > \frac{T(p_1)}{2\Omega n(p_1)}$ but $\mu < \frac{T(p_2)}{2\Omega n(p_2)}$ where $\{p_1, p_2\} \in \{0, 1\}$ with $p_1 \neq p_2$. Thus at the boundary which minimizes $T(p)/n(p)$ (p_1 boundary), mutations will be a stronger force then genetic drift, pushing the distribution towards the interior, whereas at the other boundary (p_2 boundary) the distribution will accumulate at the boundary in a state of quasi-fixation as the force of genetic drift outweighs mutations. This will tend to give rise to the sideways S-distribution as seen in Fig 1**e**,**f** in the main text.

This behaviour at the boundary prevents us from formulating a clear criteria about which type is 'favoured'. We can no longer focus exclusively upon the ratio $T(p)/n(p)$ as in the case of low mutation rate, but we also cannot use an integral measure to determine which type is stochastically favoured (i.e., does more of the mass of $\pi(p)$ occur for $p > 1/2$ or $p < 1/2$?), since $\pi(p)$ is not normalizable. However, because away from the boundaries the strongest effect upon the shape of $\pi(p)$ will be how the ratio $T(p)/n(p)$ changes in p (term (2) from eq S17), we may be inclined to argue that given sufficient 'segregating variation' exists, and so for polymorphic populations, the type minimizing the ratio $T(p)/n(p)$ will be favoured.

3. **High mutation rate**. Suppose mutation rate is sufficiently high such that the distribution is normalizable. Now we can compute $\int_{p_0}^{p_1} \pi(p)dp$, and so it makes

sense to use $\int_{1/2}^1 \pi(p)dp$ to determine which type is favoured. In particular, if $\int_{1/2}^1 \pi(p)dp > 1/2$, then the social actor is favoured (this assumes we have chosen C_0 such that $\int_0^1 \pi(p)dp = 1$). There are two possibilities here, either (a) mutations are of small effect, or (b) mutations are of large effect. We consider these cases in turn.

(a) Suppose mutations are of small effect, that is, type 1 only slightly differs from type 2 in terms of the social action. Let ν denote the social trait difference between types (ν may be positive or negative). This formulation allows for both types to be social actors: for example, if the social trait is altruism, then if $\nu > 0$ type 1 is more altruistic then type 2, while if $\nu < 0$, type 1 is less altruistic then type 2. To make it clear which quantities have a dependence on ν, we will explicitly include ν as an argument in the various functions, i.e., $T(p) = T(p, \nu)$. We will also write $\pi(p, \nu) = C_0(\nu)S(p, \nu)$, where

$$S(p,\nu) = \frac{1}{\sigma^2(p,\nu)} \exp\left(2\int^p \frac{\alpha(z)}{\sigma^2(z,\nu)}dz\right),$$

and refer to $S(p, \nu)$ as the speed measure [9]. Hence, $C_0(\nu) = 1/\int_0^1 S(p, \nu)dp$, and $\frac{dC_0}{d\nu} = -C_0(\nu)^2 \int_0^1 \frac{\partial S}{\partial \nu}dp$. Note that when $\nu = 0$, both types are identical and so $T(p, 0)/n(p, 0) = T/n$ is constant with respect to p (since there are no costs), and so $\int_{1/2}^1 \pi(p, 0)dp = 1/2$. Using this fact, a Taylor expansion of $\int_{1/2}^1 \pi(p, \nu)dp$ gives

$$\begin{aligned}
\int_{1/2}^1 \pi(p,\nu)dp &\approx \int_{1/2}^1 \pi(p,0)dp + \nu\int_{1/2}^1 \left.\frac{\partial \pi}{\partial \nu}\right|_{\nu=0} dp + \mathcal{O}(\nu^2) \\
&= \frac{1}{2} + \nu\int_{1/2}^1 \left[\frac{\partial S}{\partial \nu}C_0(\nu) + S(p,\nu)\frac{dC_0}{d\nu}\right]_{\nu=0} dp + \mathcal{O}(\nu^2) \\
&= \frac{1}{2} + \nu C_0(0)\int_{1/2}^1 \left[\frac{\partial S}{\partial \nu} - \pi(p,\nu)\int_0^1 \frac{\partial S}{\partial \nu}dz\right]_{\nu=0} dp + \mathcal{O}(\nu^2) \\
&= \frac{1}{2} + \frac{\nu C_0(0)}{2}\left[\int_{1/2}^1 \frac{\partial S}{\partial \nu}dp - \int_0^{1/2} \frac{\partial S}{\partial \nu}dp\right]_{\nu=0} + \mathcal{O}(\nu^2).
\end{aligned} \tag{S18}$$

The logic of eq S18 is clear: for example, if $\partial S/\partial \nu > 0$, then if the increase in the speed measure on the interval $(1/2, 1)$ exceeds the increase on the interval $(0, 1/2)$, we should expect the social actor to be favoured in the sense that $\int_{1/2}^1 \pi(p, \nu)dp > 1/2$ since the process will tend to spend more time in the region $(1/2, 1)$.

To make further progress, we need to compute $\left.\frac{\partial S}{\partial \nu}\right|_{\nu=0}$. For ease of notation, let $R(p, \nu) \equiv T(p, \nu)/n(p, \nu)$ and $\omega \equiv \frac{2\Omega\mu}{R}$. Then

$$\frac{\partial S}{\partial \nu} = -S(p,\nu)\left[\frac{1}{R(p,z)}\frac{\partial R}{\partial \nu} + 2\Omega\mu\int^p \frac{1-2z}{z(1-z)R(z,\nu)^2}\frac{\partial R}{\partial \nu}dz\right] \tag{S19}$$

Now what is $\partial R/\partial \nu$? Suppose we can write per-capita growth as $b(n) - d(n) + \nu\theta(n)np$, where $b(n)$ and $d(n)$ are the birth and death rates when $\nu = 0$, and $\theta(n)\nu$ is the (additional) social effect of type 1 individuals (which is multiplied by the density of type 1 individuals, np). The precise action of the social trait may be upon either the birth or death rate, while if $\theta(n)\nu < 0$, the trait

is spite whereas if $\theta(n)\nu > 0$, the trait is altruism. The function $\theta(n)$ controls any density-dependent effects of how the social action is distributed among members of the population.

Since on the slow timescale the process is in demographic equilibrium, per-capita growth is zero, and so at equilibrium

$$\nu p = \frac{b(n) - d(n)}{\theta(n) n}.$$

Let $G(n) \equiv [b(n) - d(n)]/[\theta(n)n]$, and assuming $G(n)$ is invertible, $n = g(\nu p) \equiv G^{-1}(\nu p)$. Thus we see that in order for $G(n)$ to be invertible, n must be a monotonic function of νp (which it is by our classification of the social traits). Since at equilibrium, $T(p, \nu)$ is either two times the per-capita birth rate or two times the per-capita death rate (e.g., if the social trait acts on birth rate, then the birth rate is $b(n)+\nu\theta(n)np$ and the death rate is $d(n)$ and thus $T(p, \nu) = 2(b(n) + \nu\theta(n)) = 2d(n))$, we can write R as a function of n, that is $R(n) = R(g(\nu p))$ (e.g., using our previous example, $R(n) = 2d(n)/n$). So all of the instances of ν in R are mediated through their presence in n. Then

$$\frac{\partial R}{\partial \nu} = \frac{\partial}{\partial \nu} R(g(\nu p)) = R'(n) g'(\nu p) p,$$

and when $\nu = 0$,

$$\left.\frac{\partial R}{\partial \nu}\right|_{\nu=0} = R'(g(0)) g'(0) p = \delta p$$

where $\delta \equiv R'(g(0))g'(0)$ is constant with respect to p (and ν).

Using this information and the fact that

$$S(p, 0) = \frac{\Omega}{R} \frac{1}{p(1-p)} \exp\left(\omega \int^p \frac{1-2z}{z(1-z)} dz\right) = \frac{\Omega}{R} \left(p[1-p]\right)^{\omega-1}$$

in eq S19 gives

$$\begin{aligned}
\left.\frac{\partial S}{\partial \nu}\right|_{\nu=0} &= -\frac{S(p,0)}{R}\left(p + \omega \int^p \frac{1-2z}{z(1-z)} z dz\right)\delta \\
&= -\frac{S(p,0)}{R}\left(p + \omega\left[2p + \ln(1-p)\right]\right)\delta \\
&= -\frac{\Omega}{R^2}\left([1+2\omega]p + \omega \ln(1-p)\right)\left(p[1-p]\right)^{\omega-1}\delta. \qquad \text{(S20)}
\end{aligned}$$

Now using eq S20 in eq S18 yields

$$\int_{1/2}^1 \pi(p,\nu) dp \approx \frac{1}{2} - \frac{\nu C_0(0)}{2} \frac{\Omega\delta}{R^2}\left(\int_{1/2}^1 f(p,\omega) dp - \int_0^{1/2} f(p,\omega) dp\right) + \mathcal{O}(\nu^2) \qquad \text{(S21)}$$

where

$$f(p,\omega) \equiv \left([1+2\omega]p + \omega \ln(1-p)\right)\left(p[1-p]\right)^{\omega-1}.$$

Since $\int_{1/2}^1 f dp - \int_0^{1/2} f dp$ is a function of a single variable, ω, it can be plotted. From inspection of Fig A, this quantity is positive for $\omega > 0$ and goes to zero as $\omega \to \infty$. It follows that if $\delta\nu < 0$, type 1 is favoured, whereas if $\delta\nu > 0$, type 2 is favoured. But the sign of $\delta\nu$ has the same interpretation as before: whichever type minimizes the ratio T/n is favoured. Thus when mutations are of small effect, whichever type minimizes the ratio T/n is favoured.

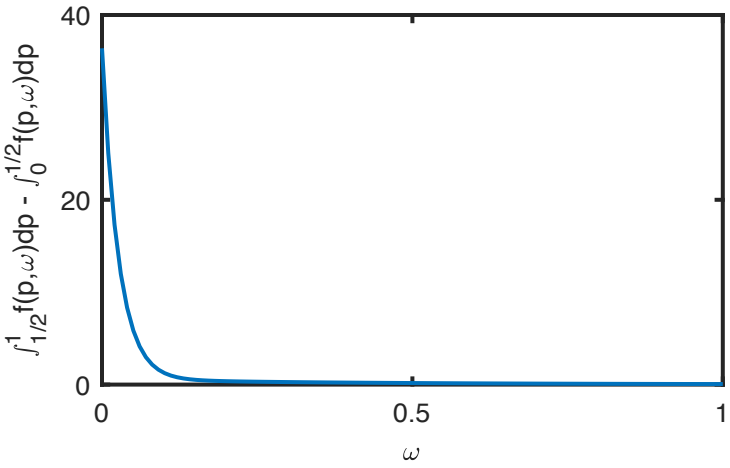

Fig A. As $\omega \to \infty$, $\int_{1/2}^{1} f dp - \int_{0}^{1/2} f dp \to 0$, but remains positive.

(b) If mutations are not of small effect, that is, $\nu \gg 0$, and mutation rate is also high, then mutations will exert a strong effect near the boundary pushing the weight of the distribution towards $p = 1/2$. In this regime, the ratio $T(p)/n(p)$ plays the strongest role. So provided $T(p)/n(p)$ is monotonic in p, then the type minimizing the ratio $T(p)/n(p)$ will tend to be favoured in the sense that the weight of the distribution will be displaced away from $1/2$ towards this boundary.

The preceding analysis assumed that the ratio $T(p)/n(p)$ was monotonic in p and so minimized at either $p = 0$ or $p = 1$. Suppose instead that $T(p)/n(p)$ is non-monotonic in p. Then for some $p \in (0, 1)$, say p^*, $\frac{d}{dp}\left[\frac{T(p)}{n(p)}\right] = 0$. From our analysis above, it is apparent where the issues are going to arise. For low mutation rate, when we are in a small neighbourhood of p^* consideration of eq S12 predicts that whichever type is more abundant will be favoured. Likewise, at intermediate mutation rates, now term (2) in eq S17 will be zero at p^* and so in a neighbourhood of p^* the dominant force shaping $\pi(p)$ will be term (1). But away from the boundary, whether term (1) favours or disfavours the social actor will depend both upon the value of p^* (is it greater or less than $1/2$?) and the magnitude of mutation rate relative to genetic drift at p^* (is μ greater or less than $T(p^*)/[2\Omega n(p^*)]$?). Finally, consider the case in which mutation rate is sufficiently high such that there is a normalizable distribution. If mutations are of small effect, that is, the two types are sufficiently similar in terms of the social trait, then implicit within the preceding analysis was that one of the types will minimize the ratio $T(p)/n(p)$ (and so $T(p)/n(p)$ is monotonic in p for sufficiently small ν). However, when mutations are of large effect, then consideration of eq S17 reveals again how the results become more complex and how the magnitude of mutation rate will alter the expected outcome.

Evolution of costly social traits

If we instead suppose that $\epsilon > 0$, then the social actor is selected against, and so can only be favoured if the influence of demographic stochasticity outweighs that of selection. This is similar to the observation from classical population genetics that in large populations, selection dominates, whereas in small populations, genetic drift does. Because of the inherent complexities owing to how the costs of the social traits are formulated (are

they density-dependent, or is $c(p)$ constant for all p?), as well as how the costs interact with mutations and demographic stochasticity, in the main text we simply focus upon numerical calculations to show that a stochastic reversal of selection is possible using the measure $\int_{1/2}^{1} \pi(p)dp$. If $\int_{1/2}^{1} \pi(p)dp > 1/2$, then despite the social actor being selected against, the population is most likely to be observed in a state in which the social actor is at greater frequency, and so we characterize this as a stochastic reversal of selection.

Examples

In this section we detail the calculations used to obtain the examples found in the main text.

1. $b(\mathbf{x}) \equiv \beta(1 - x_1 - x_2)$, $m(\mathbf{x}) \equiv d(1 + \nu x_1/[x_1 + x_2])$, and $c(\mathbf{x}) \equiv 0$ with $\beta > d(1 + |\nu|)$ and $\nu \in (-1, 1)$. Here the social action alters death rate: if $\nu < 0$ the trait is altruism, whereas if $\nu > 0$ the trait is spite. On the slow manifold, $w + \gamma_2(w) = (\beta - d(1 + \nu p))/\beta$, and so the ratio $T(p)/n(p)$ is

$$\frac{T(p)}{n(p)} = \frac{2\beta d(1 + \nu p)}{\beta - d(1 + \nu p)} \tag{S22}$$

and thus

$$\frac{d}{dp}\left[\frac{T(p)}{n(p)}\right] = \frac{2\beta^2 d\nu}{(\beta - d(1 + \nu p))^2}, \tag{S23}$$

which shares the same sign as ν. Thus if $\nu > 0$ then $T(p)/n(p)$ is increasing in p and so spite is disfavoured, whereas if $\nu < 0$ then $T(p)/n(p)$ is decreasing in p and so altruism is favoured. Simulations indicate these results extend to the n-type model (Fig 3).

2. $b(\mathbf{x}) \equiv \beta + \nu x_1$, $m(\mathbf{x}) \equiv d + \kappa_1(x_1 + x_2) + \kappa_2(x_1 + x_2)^2$ and $c(\mathbf{x}) \equiv 0$ with $\beta > d$ and $\nu > 0$. Here the social trait is altruism which alters birth rate. On the slow manifold, $\gamma_2(w) = (-2\kappa_2 w - \kappa_1 + \sqrt{4\kappa_2(\nu w + \beta - d) + \kappa_1^2})/(2\kappa_2)$, and so $h(p) = p(\nu p - \kappa_1 + \sqrt{(\nu p - \kappa_1)^2 + 4\kappa_2(\beta - d)})/(2\kappa_2)$. For this example, $T(p)/n(p)$ is non-monotonic in p (and ν) and so whether altruism is favoured or disfavoured depends upon the demographic parameters. In particular, the level of altruism minimizing $T(1)/n(1)$ is $\nu^* = (d\kappa_1 + \sqrt{\kappa_2 d(\beta - 2d)^2})/d$. Simulations predict this level of altruism also tends to be favoured in the n-type model (see Fig 3).

3. $b(\mathbf{x}) \equiv \beta(1 - \nu x_1/(x_1 + x_2 + a))$, $m(\mathbf{x}) \equiv d + \kappa_1(x_1 + x_2) + \kappa_2(x_1 + x_2)^2$ and $c(\mathbf{x}) \equiv 0$ with $\beta > d$, $a > 0$, and $\nu \in [0, 1]$. Here the social trait is spite which alters birth rate. In particular, with probability $\nu x_1/(x_1 + x_2 + a)$ a type 1 individual blocks another individual from reproducing. For this example, $\gamma_2(w)$ can be computed analytically, but the expression is unwieldy and so we do not show it here. Importantly, however, $T(p)/n(p)$ is a nonlinear function of ν, and the level of spite minimizing $T(1)/n(1)$ is

$$\nu^* = \left(a + \sqrt{\frac{d}{\kappa_2}}\right)\left(\beta - 2d - \kappa_1\sqrt{\frac{d}{\kappa_2}}\right)\sqrt{\frac{\kappa_2}{\beta^2 d}}.$$

Simulations predict this level of spite also tends to be favoured in the n-type model (Fig 3).

4. $b(\mathbf{x}) \equiv r + \nu x_1$, $m(\mathbf{x}) \equiv \kappa(x_1 + x_2)$, and $c(\mathbf{x}) \equiv r$, with $r, \nu, \kappa > 0$ and $\kappa > \nu$. Here the social trait is altruism which increases birth rate. On the slow manifold,

$\gamma_2(w) = (r - [\kappa - \nu]w)/\kappa$, and so $h(p) = rp/(\kappa - \nu p)$. Using this information yields the stationary distribution

$$\pi(p) \propto p^{\mu\Omega/\kappa - 1}(1-p)^{\mu\Omega/\kappa - 1} e^{-\epsilon r \Omega p/\kappa}, \tag{S24}$$

where proportionality is up to a positive constant. In the absence of costs ($\epsilon = 0$), eq S24 is symmetric about $p = 1/2$, and so neither type is stochastically favoured. The reason for this result is that $T(p)/n(p) = 2\kappa$ which is constant for all p.

5. $b(\mathbf{x}) \equiv \beta + \nu x_1$, $m(\mathbf{x}) \equiv d + \kappa(x_1 + x_2)$, and $c(\mathbf{x}) \equiv r$, with $\beta > d > 0$, $\kappa > \nu > 0$ and $r = \beta - d$, and so the social trait is altruism increasing the birth rate. The per-capita growth rate in this model is the same as in example 4 and so $\gamma_2(w)$ and $h(p)$ are the same. However now the stationary distribution is

$$\pi(p) \propto p^{\frac{\mu\Omega r}{\beta\kappa} - 1}(1-p)^{\frac{\mu\Omega r}{\beta\kappa - d\nu} - 1}(\beta\kappa - d\nu p)^{\frac{r^2\Omega\epsilon}{d\nu} - \frac{\mu\Omega r}{\beta\kappa} - \frac{\mu\Omega r}{\beta\kappa - d\nu} - 1}.$$

Here, in the absence of costs, the distribution is asymmetric and favours type 1 (the altruist). This can be seen by noticing that $T(p)/n(p) = 2(\beta\kappa - d\nu p)/r$, which is decreasing in p.

Interestingly, example 4 is the non-spatial model of Constable, Rogers, McKane & Tarnita [8] (CRMT), differing only in that we have explicitly included mutations (our notation also slightly differs). CRMT concluded that stochasticity induced an advantage for the altruist whereas our analysis shows altruism is stochastically neutral if cost-free and disfavoured otherwise. There are two reasons for this discrepancy.

1. Density versus frequency. Rather than dealing with the SDE for proportions directly (eq S9), CRMT focused upon interpreting the infinitesimal mean of the density SDE, dx_1 (this can be obtained from (S9) by application of Ito's formula [1]). However, evolution is a change in frequency, not density, and a change in density is not equivalent to a change in frequency. This is clear from eq S9 where in the absence of mutations and selection, the infinitesimal mean is zero while $\sigma^2(p) = \sqrt{\frac{2\kappa}{\Omega}p(1-p)}$ so the stochastic process is mathematically equivalent to pure genetic drift in a population of constant size [14].

2. Inclusion of mutations. At selective neutrality, which is the scenario most conducive to the evolution of altruism, the fixation probability of a particular type is equal to its proportion in the population. Thus the invasion probability of a type j mutant into a monomorphic type i population of size N_i is simply $1/N_i$. As the altruist can grow to a larger population size then the non-altruist, pairwise comparison of invasion probabilities predicts the altruist is favoured [8, 15, 16]. However, if we assume type i mutates to type j at a per-capita rate μ_{ij}, then as shown in eq S13 the transition rate from an all-type i state to an all-type j state is μ_{ij}. So in the absence of mutational biases, the Markov chain is equally likely to be in any state, a standard result for neutral evolution in sequential-fixation models [11–13]. Hence mutations erase any numerical advantage of the altruists. However, this does not take into account that in our model the expected time till fixation varies based upon population composition. Consideration of expected time till fixation (or absorption time) reveals the importance of the ratio $T(p)/n(p)$ (see eq S16).

Simulations

To support our analytic predictions, we use two types of simulations: Gillespie's algorithm [17] and the Euler-Maruyama (EM) method [3]. In particular, for plots involving the

stationary distribution of the two-type model, we have used Gillespie's algorithm (Fig 1, Fig 2 and Fig 4) to simulate the full stochastic process specified by eq S1. When we consider more then two types (as in Fig 3) we simulate the system of stochastic differential equations which approximates eq S1 using the EM method [3], which we detail briefly here. For the EM method, we assume that a type i individual mutates to a type j individual at rate μ. Thus the total rate at which a type i individual mutates to a *different* type is $\mu(n-1)$. Let η be a $n \times n$ matrix of standard normal random variables, and let $\Delta\tau$ be the step-size. Then using EM, the change in variable x_i over the time increment $\Delta\tau$, i.e., $\Delta x_i \equiv x_i(\tau + \Delta\tau) - x_i(\tau)$, is given by

$$\begin{aligned}\Delta x_i = \big((b_i(\mathbf{x}) - m_i(\mathbf{x}) - \mu n)x_i + \sum_j \mu x_j\big)\Delta\tau \\ + \sqrt{\frac{(b_i(\mathbf{x}) + m_i(\mathbf{x}))x_i\Delta\tau}{\Omega}}\eta_{ii} + \sum_j \left(\sqrt{\frac{\mu x_j \Delta\tau}{\Omega}}\eta_{ij} - \sqrt{\frac{\mu x_i \Delta\tau}{\Omega}}\eta_{ji}\right).\end{aligned} \tag{S25}$$

For all the n-type simulations used in this paper, $\Delta\tau = 0.01$, $\Omega = 10^4$, and $\mu = 10^{-6}$, and the initial conditions were chosen to be $x_i(0) = 0.05$ for all i; this was then simulated until $\tau = 2 \times 10^6$ for 10^4 individual simulations (we checked that the distribution had settled down by $\tau = 2 \times 10^6$). Then if $x_i^{(j)}(\tilde{\tau})$ is the density of type i in sample run j at time $\tilde{\tau}$, then the probability of type i plotted in Fig 3 (black circles) is

$$\text{Prob}(x_i) = \sum_j \frac{x_i^{(j)}(\tilde{\tau})}{\sum_k x_k^{(j)}(\tilde{\tau})} \Bigg/ \left[\sum_\ell \sum_j \frac{x_\ell^{(j)}(\tilde{\tau})}{\sum_k x_k^{(j)}(\tilde{\tau})}\right]. \tag{S26}$$

The primary reason for using EM method rather than Gillespie's algorithm when the number of types is greater then 2 is that the computational costs of Gillespie's algorithm rapidly become prohibitive. This is because when we construct the stationary distribution in the 2-type case, in order to have the distribution normalizable, mutations must be artificially high. When mutations are high, the population composition can change more rapidly, and so simulations reach the stationary distribution more rapidly. When we extend the system to include more than 2-types, mutations have a homogenizing effect, and so we lower the mutation rate. However, this means that the population composition changes less quickly, and so more time must elapse to obtain the stationary distribution.

References

1. Gardiner CW. Handbook of Stochastic Methods. Berlin, Germany: Springer: Complexity; 2009.
2. van Kampen NG. Stochastic Processes in Physics and Chemistry. New York, NY: North-Holland; 1981.
3. Allen LJS. An Introduction to Stochastic Processes with Applications to Biology. Boca Raton, FL: CRC Press; 2011.
4. Carr J. Applications of Centre Manifold Theory. New York, NY: Springer-Verlag; 1981.
5. Berglund N, Gentz B. Noise-induced Phenomena in Slow-Fast Dynamical Systems. USA: Springer; 2006.

6. Parsons TL, Rogers T. Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: a constructive approach with examples from theoretical biology. J Phys A: Mathematical and Theoretical. 2017;50.

7. Constable GWA, McKane AJ, Rogers T. Stochastic dynamics on slow manifolds. J Phys A: Math Theor. 2013;46.

8. Constable GWA, Rogers T, McKane AJ, Tarnita CE. Demographic noise can reverse the direction of deterministic selection. Proc Nat Acad Sci. 2016;.

9. Karlin S, Taylor HM. A Second Course in Stochastic Processes. New York, NY: Academic Press; 1981.

10. Ewens WJ. Mathematical Population Genetics I: Theoretical Introduction. New York, NY: Springer: Interdisciplinary Applied Mathematics; 2004.

11. McCandlish DM, Stoltzfus A. Modeling evolution using the probability of fixation: history and implications. Quarterly Review of Biology. 2014;89:225–252.

12. Kimura M. Evolutionary rate at the molecular level. Nature. 1968;217:624–626.

13. King JL, Jukes TH. Non-Darwinian evolution. Science. 1969;164:788–798.

14. Kimura M. Solution of a process of random genetic drift with a continuous model. Proc Natl Acad Sci. 1955;41:144–150.

15. Houchmandzadeh B, Vallade M. Selection for altruism through random drift in variable size populations. BMC Evol Biol. 2012;12.

16. Houchmandzadeh B. Fluctuation driven fixation of cooperative behavior. Biosystems. 2015;127:60–66.

17. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem. 1977;81:2340–2361.