PUBLIC GOODS GAMES IN POPULATIONS WITH FLUCTUATING SIZE

ALEX MCAVOY[1], NICOLAS FRAIMAN[2], CHRISTOPH HAUERT[3], JOHN WAKELEY[4], AND
MARTIN A. NOWAK[1,4,5]

[1] *Program for Evolutionary Dynamics, Harvard University, Cambridge, MA 02138*
[2] *Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599*
[3] *Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, BC, Canada V6T 1Z2*
[4] *Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138*
[5] *Department of Mathematics, Harvard University, Cambridge, MA 02138*

ABSTRACT. In social dilemmas, cooperation can be essential to the existence of a population when defectors lack the ability to survive on their own. A model of this phenomenon must incorporate stochastic fluctuations in population size to allow for extinction. Here, we present a game-theoretic extension of the Wright-Fisher process that accounts for demographic stochasticity. A population of defectors (trivially) always survives when the population size is fixed, but, if extinction is possible, then their survival can depend on the presence of cooperators. When the population is small and polymorphic, the effects of random drift are initially strong and can result in the immediate extinction of defectors, which allows cooperators to thrive in the absence of exploitation by defectors. Moreover, allowing dynamic population size results in a more multifaceted notion of a trait's evolutionary success relative to models with constant size. In particular, we show that selection can decrease the relative frequency of cooperators while increasing both their absolute abundance and the overall population size, a property that is unique to evolving populations with non-constant population size.

1. INTRODUCTION

The emergence of cooperation is a central topic of study in evolutionary theory. The way the problem is usually formulated, it pays to exploit cooperators, yet the payoff to one cooperator against another is greater than the payoff to one defector against another [1]. In spite of this dilemma, cooperation is broadly observed in nature, and various mechanisms have been put forth to explain its evolution [2, 3]. The question of how cooperators may proliferate in social situations is one of the main concerns of evolutionary game theory, in which cooperation and defection are modeled as strategies of a game.

Evolutionary game-theoretic models typically involve a number of assumptions. Here, we are concerned with two potentially restrictive ones: (i) the population size is infinite or (ii) the population size is finite but fixed and unaffected by evolution. While the classical replicator equation of Taylor and Jonker [4] can be used to model large populations that fluctuate in size [5], the dynamics themselves capture only the relative frequencies of the strategies. Even stochastic models that account for populations of any finite size, such as the Moran model or the Wright-Fisher model and their game-theoretic extensions, typically require the number of players to remain static [6–16]. Here, we explore the evolutionary dynamics of cooperation in social dilemmas when the population size is finite and can fluctuate.

Although branching processes that model populations of variable size have a rich history in theoretical biology [see 17], it is only recently that stochastic models with fluctuating population size have been considered within the realm of evolutionary game theory. Hauert et al. [5] treat ecological dynamics in evolutionary games by modifying the replicator equation to account for changing population density, and show that fluctuating density can lead to coexistence between cooperators and defectors. Melbinger et al. [18] illustrate how the decoupling of stochastic birth and death events can lead to transient increases in cooperation. By allowing a game to influence carrying capacities, Novak et al. [19] demonstrate that variable density regulations can change the stability of equilibria compared to the classical replicator equation. Furthermore,

demographic fluctuations can act as a mechanism to promote cooperation in public goods games [20] and indefinite coexistence (as opposed to fixation) in coexistence games [21]. Fluctuating size in a generalized Lotka-Volterra model also leads to different growth rates for isolated populations of cooperators and defectors [22]. Even when traits have the same growth rate, these fluctuations can still affect a mutant's fixation probability [23]. On the whole, demographic stochasticity has received relatively little attention in evolutionary game theory, but its effects are already observed to be extremely important for evolving populations.

We propose an extension of Wright-Fisher [24, 25] evolutionary games that accounts for populations that fluctuate in size. The Wright-Fisher process consists of non-overlapping generations in which trait values of offspring are sampled from the previous generation based on the success of individuals (parents) with those traits in a sequence of interactions [26, 27]. The payoffs from these interactions are determined by a game and represent the competition between the different types. However, the process is usually defined such that every subsequent generation has exactly the same size as the first generation. We consider the implications of removing this assumption for the evolution of cooperation.

When a population's size varies, both extinction and unbounded growth become possibilities. In order to understand the evolution of cooperation, we first need to model the ecological dynamics of the population. We capture these dynamics in a single quantity, f_N, which represents the per-capita reproductive rate when the population size is N. We assume that this rate is a non-increasing function of the population size, N, so that larger populations lead (in general) to lower per-capita reproductive rates due to ecological constraints. This baseline rate is the fluctuating-size analogue of the "background fitness" that is typically used in models with fixed population size [14, 28].

If there is a population size, N^*, for which $f_{N*} = 1$, then each individual produces (on average) one offspring, so the population size is expected to remain constant. Such an N^* is thus a carrying capacity for the population. Many traditional models justify the assumption of fixed population size by assuming that the population has reached its carrying capacity [29] and will never go extinct. In these models, therefore, there is an implicit assumption that population regulation is extremely strong. Moreover, changes to the composition of a population (such as through the proliferation of cooperation) do not affect this carrying capacity. Our aim here is to show how traits like cooperation affect the background ecological dynamics defined by f_N. In particular, we explore the influence of cooperation on "metastable" states that occur when the population fluctuates about its carrying capacity and show how the carrying capacity itself depends on the parameters of the model.

We consider a public goods game in which cooperators put effort toward a public resource at a cost, w, to their fecundity. Defectors do not put forth this effort and so do not incur a cost. The public resource is distributed evenly among all individuals in the population, but its per-capita effect on fecundity can be greater than the per-capita cost of its production [30]. We use $R \geqslant 1$ to denote the multiplication factor of the public good, which quantifies the return on investment in the public good. Incidentally, the terms "cooperate" and "defect" in this setting refer to the fact that when the group (population) cooperates, everyone is better off than if everyone defected due to the multiplication factor, R; nonetheless, defectors can free-ride off of cooperators and benefit from the public good while incurring none of the costs of maintaining it. Note that w quantifies the selective difference between cooperators and defectors. When $w = 0$, the two types are indistinguishable based on fecundity.

Public goods games have been used to model conflicts of interest ranging from cooperation in microorganisms [31–35], to alarm calls in monkeys [36, 37], sentinel behavior in meerkats [38], and large-scale human efforts aimed at combating climate change [39, 40] and pollution [41]. Due to its linearity and close relationship to the Prisoner's Dilemma, the public goods game we consider is sometimes called the "N-person Prisoner's Dilemma" [42]. Provided $1 \leqslant R \leqslant N$, this game presents a conflict of interest between the group and the individual that can be reduced to a sequence of $N-1$ Prisoner's Dilemma interactions [43]. However, the analysis and interpretation of a single public goods game is somewhat more straightforward than that of a series of Prisoner's Dilemma interactions when the population size fluctuates.

When the population size is fixed, selection affects population dynamics through its effects on relative reproductive rates. That is, the selective advantage of one trait is measured relative to another trait and the baseline reproductive rate, f_N, is irrelevant. In fluctuating populations, it is more natural to interpret the reproductive rate of a player in absolute terms as that player's expected number of surviving offspring. (The expected number of offspring is sometimes called "absolute fitness," but here we favor the term "reproductive

rate" since it suggests a more mechanistic interpretation [see 44].) From an evolutionary perspective, an (absolute) reproductive rate then affects both the relative frequencies of traits as well as their total quantities.

We are particularly interested in the case when a population of defectors cannot survive on their own but a population of cooperators can. Since a population of defectors evolve according to the baseline reproductive rate, f_N, and since f_N is non-increasing here, they can grow (on average) to a non-zero carrying capacity only when $f_1 \geqslant 1$. A population of cooperators, on the other hand, evolves according to the reproductive rate $(1 + w(R-1)) f_N$, which can be greater than 1 even when $f_N < 1$. In polymorphic populations, such as when offspring can mutate from the parental type, cooperators and defectors have reproductive rates $F_C = F_C(x_C, x_D)$ and $F_D = F_D(x_C, x_D)$, where x_C and x_D represent the number of cooperators and defectors in the population, respectively, which depend on the baseline rate, f_N, the cost of cooperation, w, and the multiplication factor for the public good, R.

We show here that when the population size can fluctuate, selection can be essential for the survival of the population as a whole. Thus, even when cooperators are less frequent than defectors in the mutation-selection equilibrium, there can be an optimal cost of cooperation, w^*, depending on the baseline reproductive rate, f_N, at which (i) the population does not immediately go extinct, with the numbers of cooperators and defectors each fluctuating around equilibrium values, and (ii) the frequency of cooperators is maximized subject to (i). In other words, cooperation can be favored by selection at a positive cost of cooperation when there is demographic stochasticity, which marks a departure from the behavior of models with fixed size.

Furthermore, even when the population can survive under neutral drift, selection can increase the number of cooperators despite decreasing their frequency. If the population size is fixed, then cooperators are less frequent than defectors if and only if cooperators are less abundant than defectors. However, this equivalence breaks down when the population size can fluctuate because the frequency of a strategy is determined by both its abundance and the population size. Therefore, the evolutionary success of a strategic type depends on more than just the strategy; it also depends on the dynamics of the underlying neutral process.

2. Description of the model

2.1. **Update rule.** In the Wright-Fisher process, the population has fixed size, N. We assume that individuals reproduce asexually. Although we do not model genetic transmission explicitly, our model corresponds to one in which individuals are haploid. In a game with two strategies, C ("cooperate") and D ("defect"), it follows that the state of the population is determined by number of cooperators, which we denote by x_C. If $F_C = F_C(x_C)$ and $F_D = F_D(x_C)$ give the (frequency-dependent) reproductive rates of cooperators and defectors, respectively, in the state with x_C cooperators, then the probability of transitioning from this state to the state with y_C cooperators (provided $0 \leqslant y_C \leqslant N$) is

$$\mathbf{P}(x_C, y_C) = \binom{N}{y_C} \left(\frac{x_C F_C}{x_C F_C + (N - x_C) F_D} \right)^{y_C} \left(\frac{(N - x_C) F_D}{x_C F_C + (N - x_C) F_D} \right)^{N - y_C}. \tag{1}$$

In other words, the cooperators in one generation are sampled from the previous generation according to the binomial distribution with mean $N x_C F_C / (x_C F_C + (N - x_C) F_D)$. One biological interpretation for this transition rule is the following: Each player in one generation produces a large number of gametes from which the surviving offspring in the next generation are selected. These offspring are sampled at random, weighted by the success of the parents in competitive interactions modeled by the game, subject to a constant population size.

Extending the Wright-Fisher process to populations that fluctuate in size involves simply removing the condition that the next generation has the same size as the current generation. Suppose that $F_C = F_C(x_C, x_D)$ and $F_D = F_D(x_C, x_D)$ give the reproductive rates of cooperators and defectors, respectively, in state (x_C, x_D). We assume that the number of offspring of each individual follows a Poisson distribution, with parameter F_C if the individual is a cooperator and parameter F_D if the individual is a defector. Then, the probability of transitioning from state (x_C, x_D) to state (y_C, y_D) is

$$\mathbf{P}\Big((x_C, x_D), (y_C, y_D) \Big) = \left(\frac{(x_C F_C)^{y_C} e^{-x_C F_C}}{y_C!} \right) \left(\frac{(x_D F_D)^{y_D} e^{-x_D F_D}}{y_D!} \right), \tag{2}$$

which reduces to **Eq. 1** when the process is conditioned on having constant population size (see **Supporting Information**). Similar branching processes have been treated elsewhere [see 45], but, to our knowledge, a Wright-Fisher branching process has not been considered within the context of evolutionary game theory.

The transition probabilities of **Eqs. 1–2** do not take into account errors in strategy transmission, i.e. mutations. In what follows, we assume that when an individual reproduces, the offspring acquires a random strategy with probability $u \geqslant 0$. Thus, with probability $1 - u$, the offspring acquires the strategy of the parent and with probability u, becomes either a cooperator or a defector (uniformly at random). Using the binomial distribution with parameter q, whose probability density function is $b_q(n, k) := \binom{n}{k} q^k (1 - q)^{n-k}$, this mutation rate is incorporated into the transition rule defined by **Eq. 2** as follows:

$$\mathbf{P}^u\Big((x_C, x_D), (y_C, y_D)\Big) = \sum_{z_C=0}^{y_C+y_D} \mathbf{P}\Big((x_C, x_D), (z_C, y_C + y_D - z_C)\Big)$$

$$\times \sum_{k=\max\{0, z_C-y_C\}}^{\min\{z_C, y_D\}} b_{u/2}(z_C, k) \, b_{u/2}(y_C + y_D - z_C, k + y_C - z_C). \tag{3}$$

In words, we sum over all transitions defined by **Eq. 2** such that after mutations are accounted for, the number of cooperators is y_C and the number of defectors is y_D. Note that mutations themselves do not alter the population size.

2.2. **Public goods game.** Moving from fixed to variable population size in the Wright-Fisher model requires replacing relative reproductive rates by absolute reproductive rates. While the relative rate of reproduction determines competition between cooperators and defectors, absolute reproductive rates quantify population growth rates in addition to competition between the traits. Here, we use a public goods game to illustrate the application of dynamically adjusted reproductive rates.

As in models with fixed population size, one can use game theory to perturb the reproductive rates of players according to their strategies. When the population size is fixed and there is no selection, all strategies are indistinguishable and it suffices to assume that this reproductive rate is understood in a relative sense and equal to 1. When the population fluctuates in size, it could grow, go extinct, or stay roughly fixed in size; these behaviors are determined by the neutral rate, which is understood in an absolute sense. Therefore, the effects of a game on reproductive rates can be expected to influence both strategy frequencies and overall population size.

Specifically, in a population of size N, let f_N be the expected number of surviving offspring of any given player. While this reproductive rate does not vary from player to player, it can depend on the number of players in the population, N. If there are limited resources and reproduction slows as the population grows, then f_N is a decreasing function of N. An example is $f_N = f_K + r\left(1 - \frac{N}{K}\right)$ for some f_K, r, and K. In this case, r reflects the growth rate when the population is small and K satisfies $f_K = 1$, i.e. the population size at which each player replaces itself with one offspring, on average. f_K is a parameter describing the reproductive rate when the population has size $N = K$. To ensure f_N is non-negative, we set $f_N = 0$ whenever $f_K + r\left(1 - \frac{N}{K}\right) \leqslant 0$. Similarly, we also consider the case in which the reproductive rate is constant up to a threshold and decreasing thereafter, in which case $f_N = 1 + r$ if $N \leqslant K$ and $f_N = (1 + r)(K/N)$ if $N > K$ for some r and K. For both of these functions, $f_N \to 0$ as $N \to \infty$.

Consider a game with two strategies, C ("cooperate") and D ("defect"), and suppose that a defector does nothing and a cooperator incurs a cost, w, of his or her reproductive rate, f_N, in order to contribute to the provision of a public good. This contribution is then multiplied by a factor of $R \geqslant 1$, the "multiplication factor" of the public good, and distributed evenly among the N players in the population [30]. This multiplication factor may be interpreted as quantifying the return on investment in the common pool (i.e. public good) [46].

If $N = x_C + x_D$, then the reproductive rates of cooperators and defectors, $F_C = F_C(x_C, x_D)$ and $F_D = F_D(x_C, x_D)$, are given by

$$F_C = \left(1 - w + wR\frac{x_C}{N}\right) f_N; \tag{4a}$$
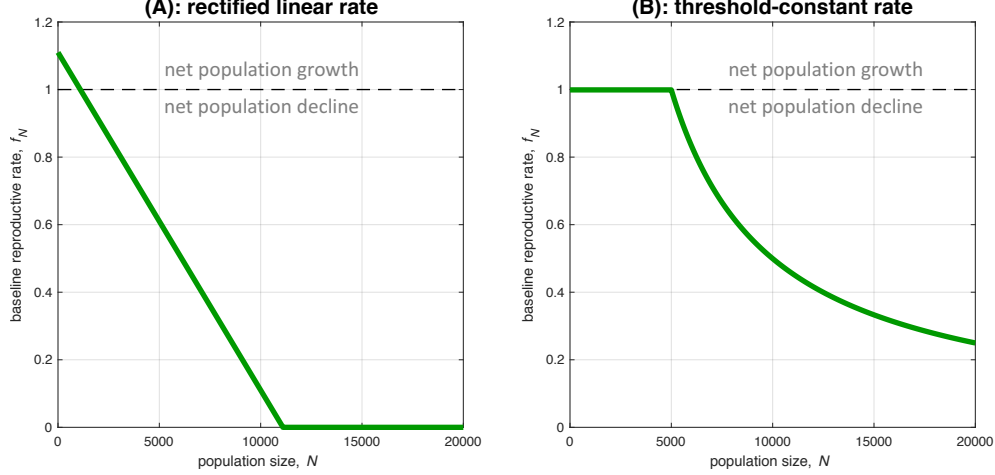
$$F_D = \left(1 + wR\frac{x_C}{N}\right) f_N. \tag{4b}$$

FIGURE 1. Two examples of baseline reproductive rates, f_N. In (A), $f_N = \max\left\{0, f_K + r\left(1 - \frac{N}{K}\right)\right\}$ for $f_K = 1.1$, $r = 0.01$, and $K = 100$. This function decreases linearly to 0, which is first achieved at $N^* = K\left(1 - \frac{1}{r}\left(1 - f_K\right)\right)$, and remains at 0 for all $N \geqslant N^*$ since the reproductive rate is (by definition) non-negative. This function is called a "rectified" linear rate because of the constraint that $f_N \geqslant 0$. In (B), $f_N = (1 + r)\min\left\{1, \frac{K}{N}\right\}$ for $r = -0.001$ and $K = 5000$. The reproductive rate is held constant and equal to $1 + r$ whenever $1 \leqslant N \leqslant K$. When the population size is greater than K, the resources contributing to reproductive rates are no longer abundant and must be divided up among the individuals in the population, which gives the expression $f_N = (1 + r)\frac{K}{N}$ for $N \geqslant K$. This function is a "threshold-constant" rate since it is constant up to the threshold population size $N = K$. Note that in (A), when the population size is initially small, there is net population growth since $f_1 > 1$. In (B), $f_N < 1$ for every N, so any population evolving according to this rate function will (on average) shrink in the direction of extinction. This immediate extinction could be prevented by choosing $r > 0$ instead of $r < 0$ as in (B).

Note that if $\pi_C\left(x_C, x_D\right) = R\left(\frac{x_C}{x_C + x_D}\right) - 1$ and $\pi_D\left(x_C, x_D\right) = R\left(\frac{x_C}{x_C + x_D}\right)$ are the classical payoff functions for the public goods game [42], then **Eq. 4a** and **Eq. 4b** correspond to cooperators and defectors having (absolute) reproductive rates $F_C = (1 + w\pi_C)f_N$ and $F_D = (1 + w\pi_D)f_N$, respectively.

When $w \ll 1$, the contribution of this game to reproductive rate is small. On the other end of the spectrum, when $w = 1$, a player cooperates by contributing the entirety of their baseline reproductive rate, f_N. Unlike in many evolutionary games in populations of fixed size, where w represents selection strength and quantifies relative differences between traits, here the cost of cooperation admits an intuitive biological interpretation: If one separates an investment in the public good from the reward obtained from it, then w represents the fraction of the expected number of surviving offspring a cooperator risks were there no return on investment.

If $f_1 < 1$, then, for a population whose dynamics are governed by a non-increasing baseline reproductive rate, f_N, there is no positive metastable capacity at which a population of defectors can sustain itself. Since generations are non-overlapping, the death of each individual in the parental generation is matched by less than one birth (on average), so the average net growth rate of the population is negative. This behavior is similar to what is seen in spatial games with empty cells, where defectors (alone) can be incapable of colonizing empty areas of the population [47]. If $f_1 \geqslant 1$, then this net growth rate is non-negative when the population is small. Therefore, since $f_N \to 0$ as $N \to \infty$, a population of defectors will grow to a stochastic carrying capacity and fluctuate around this size. (For the two classes of baseline rates we consider here (**Fig. 1**), this carrying capacity need not be exactly the parameter K; see **Supporting Information**). Therefore, defectors fall (broadly) into one of two classes: $f_1 < 1$, where no non-zero carrying capacity exists, and $f_1 \geqslant 1$, where there exists a positive carrying capacity for a population of defectors.

Note that f_1 itself does not determine the extinction probability of the (neutral) population: if $f_N = f_1 > 1$ for all $N \geqslant 1$, then the extinction probability is strictly less than 1. On the other hand, for $f_N = (1 + r) \min \left\{ 1, \frac{K}{N} \right\}$ with $r > 0$ and $K \geqslant 1$, we have $f_1 > 1$ even though the population goes extinct with probability 1. Since we consider only non-increasing rate functions, f_N, that satisfy $f_N < 1$ when N is large enough, the purpose of classifying baseline rates according to f_1 is simply to establish whether or not the neutral population (or, equivalently, a population of defectors) has a non-zero carrying capacity.

3. Evolutionary dynamics of the Wright-Fisher branching process

Here, we study how the cost of cooperation relative to f_N affects evolutionary dynamics.

3.1. **Selection dynamics (without mutation).** When the initial numbers of cooperators and defectors are small, stochastic effects have a profound influence over the long-run composition of the population. We show in **Supporting Information** that any non-zero metastable equilibrium must be monomorphic (all-cooperator or all-defector) for the update rule defined by **Eq. 2**. Although defectors generally have larger growth rates than cooperators in mixed populations, they can go extinct nearly immediately in small populations, which, in turn, can permit cooperators to prosper. For example, suppose that defectors cannot survive on their own ($f_1 < 1$), which means that any population of defectors shrinks, on average, from one generation to the next. If any population of cooperators grows, due to the multiplication factor $R > 1$, then the only populations that persist beyond a short time horizon are those composed entirely of cooperators. Therefore, cooperators have a type of survivorship bias. **Fig. 2** illustrates this phenomenon, showing that defectors often outcompete cooperators (approximately 85% of the time) when both are in the population (A), but once one type goes extinct, the population must consist of just cooperators in order to survive for a non-negligible amount of time (B). These simulations are done with the baseline rate $f_N = \max \left\{ 0, f_K + r \left(1 - \frac{N}{K} \right) \right\}$, where $f_K = 0.99$, $r = 0.01$, and $K = 100$; cost of cooperation $w = 0.1$; and multiplication factor $R = 2.0$.

3.2. **Mutation-selection dynamics.** A common way to quantify the evolutionary success of cooperators is to introduce strategy mutations and study the frequency of cooperators in the mutation-selection equilibrium [48–50]. Mutations indicate errors in the transmission (either cultural or genetic) of the two strategies (cooperation and defection) and can be small [2, 51] or large [52] depending on their interpretation. The success of cooperation is quantified by its average frequency in the population over many generations. In a population of cooperators and defectors under neutral drift (i.e. without selective differences between the two types), cooperators are indistinguishable from defectors and are equally frequent in the mutation-selection equilibrium. If selection results in a cooperator frequency above this neutral value of $1/2$, then selection is said to favor the evolution of cooperation. By this metric, selection typically disfavors cooperation in well-mixed populations [48].

If the population size is static and the update rule is that of the Wright-Fisher process, **Eq. 1**, then the baseline reproductive rate appearing in **Eq. 4**, f_N, cancels out. Only the relative fitnesses of cooperators and defectors matters. The dynamics are then captured in the relative frequencies of cooperators and defectors. Since cooperators are always less frequent than defectors when the intensity of selection, w, is positive, selection unambiguously disfavors cooperators relative to defectors. This result can be seen in **Fig. 3**(D)–(F), in which results are shown for three different values of w. These simulations were generated using a multiplication factor of $R = 2.0$ and a mutation rate of $u = 0.01$. That selection favors defectors is a standard property of many social dilemmas in well-mixed populations, where additional mechanisms–such as spatial structure–must be present for cooperators to outperform defectors [14].

When the population size can fluctuate and u is the probability that a mutation occurs, then the dynamics are governed by **Eq. 3**. Here, it is still the case that selection decreases the frequency of cooperators relative to defectors. On the other hand, the population can quickly go extinct if selection is not sufficiently strong, which we illustrate in **Fig. 3**(A)–(C) with $R = 2.0$, $u = 0.01$, and $f_N = (1 + r) \min \left\{ 1, \frac{K}{N} \right\}$ with $r = -0.001$ and $K = 5000$. Thus, cooperation can be favored in such situations because it protects against extinction.

Let $\mathbf{E}_{(x_C, x_D)} [y_C]$ (resp. $\mathbf{E}_{(x_C, x_D)} [y_D]$) denote the expected cooperator (resp. defector) abundance in the next generation given x_C cooperators and x_D defectors in the current generation. Past a selection threshold, w^*, the population survives and reaches a metastable equilibrium, which is defined as a state consisting of
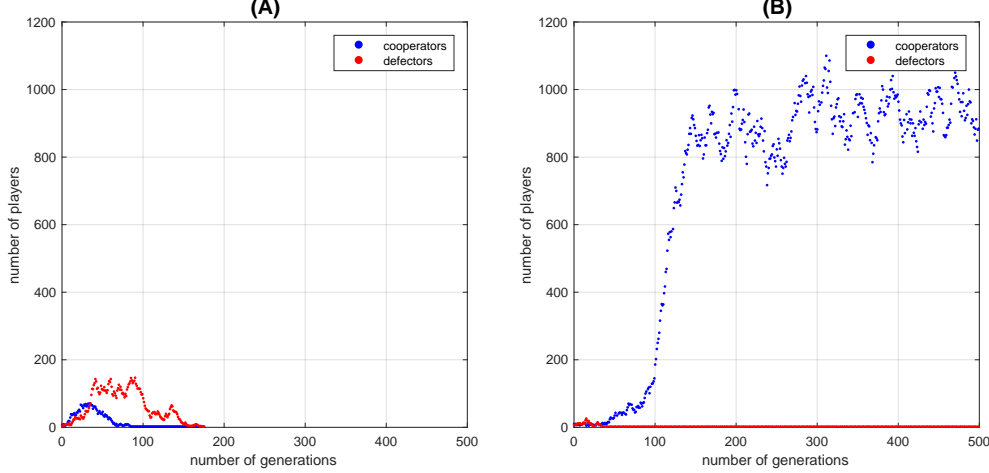
FIGURE 2. Dynamics of the Wright-Fisher branching process in the absence of mutation. The baseline reproductive rate is given by $f_N = \max\left\{0, f_K + r\left(1 - \frac{N}{K}\right)\right\}$, where $f_K = 0.99$, $r = 0.01$, and $K = 100$. The initial population consists of just 5 cooperators and 5 defectors. Approximately 15% (rounded to the nearest percentage) of runs result in the behavior of panel (B), with defectors going extinct and cooperators reaching their carrying capacity. The remaining runs resemble panel (A), with almost-immediate extinction of the entire population. We show in **Supporting Information** that, without mutation, any metastable equilibrium is necessarily monomorphic; since defectors cannot survive on their own ($f_1 < 1$), it follows that only all-cooperator states can be observed as the long-run outcome of these initial conditions. Notably, while defectors go extinct in fewer than 200 generations in (A), in (B) the population of cooperators thrives even after 10^9 generations (although eventually it will go extinct). To demonstrate the initial ascent, we include here only the first 500 generations. Parameters: $u = 0$, $w = 0.1$, and $R = 2.0$.

x_C^* cooperators and x_D^* defectors such that

$$\mathbf{E}_{\left(x_C^*, x_D^*\right)}\left[y_C\right] = x_C^*; \tag{5a}$$

$$\mathbf{E}_{\left(x_C^*, x_D^*\right)}\left[y_D\right] = x_D^*, \tag{5b}$$

If (x_C^*, x_D^*) is a metastable equilibrium and σ denotes standard deviation, then

$$\sigma_{\left(x_C^*, x_D^*\right)}\left[y_C\right] / \mathbf{E}_{\left(x_C^*, x_D^*\right)}\left[y_C\right] = \sqrt{x_C^*}; \tag{6a}$$

$$\sigma_{\left(x_C^*, x_D^*\right)}\left[y_D\right] / \mathbf{E}_{\left(x_C^*, x_D^*\right)}\left[y_D\right] = \sqrt{x_D^*}, \tag{6b}$$

which means that the fluctuations around a metastable state constitute only small fractions of x_C^* and x_D^* when x_C^* and x_D^* are large (see **Supporting Information** for further details).

One key difference from models with fixed population size is that, in a branching process, the population either grows unboundedly or else eventually goes extinct [53, 54]. That is, if the population remains bounded in size, then the only true stationary state is extinction. Despite this behavior capturing the long-run dynamics of the process, there can also exist metastable states in which the process persists prior to population extinction. We show in **Supporting Information** that this persistence time is actually quite long. In particular, the expected time to extinction grows exponentially in K [see also 55], meaning that if $\mathbf{E}\left[\tau_K\right]$ is the expected number of generations prior to extinction after starting in the metastable state, then there exists $c > 0$ (independent of K) for which $\mathbf{E}\left[\tau_K\right] \geqslant e^{cK}$. Informally, x_C^* and x_D^* grow with K, and since the fluctuations in x_C^* and x_D^* are on the order of $\sqrt{x_C^*}$ and $\sqrt{x_D^*}$, respectively, the amount of time until deviations from the mean destroy the population grows rapidly in K (see **Fig. 4**). **Fig. 3**(C) illustrates persistence of the population at a metastable state prior to (eventual) extinction.

7

Dynamic population size (Wright-Fisher branching process)

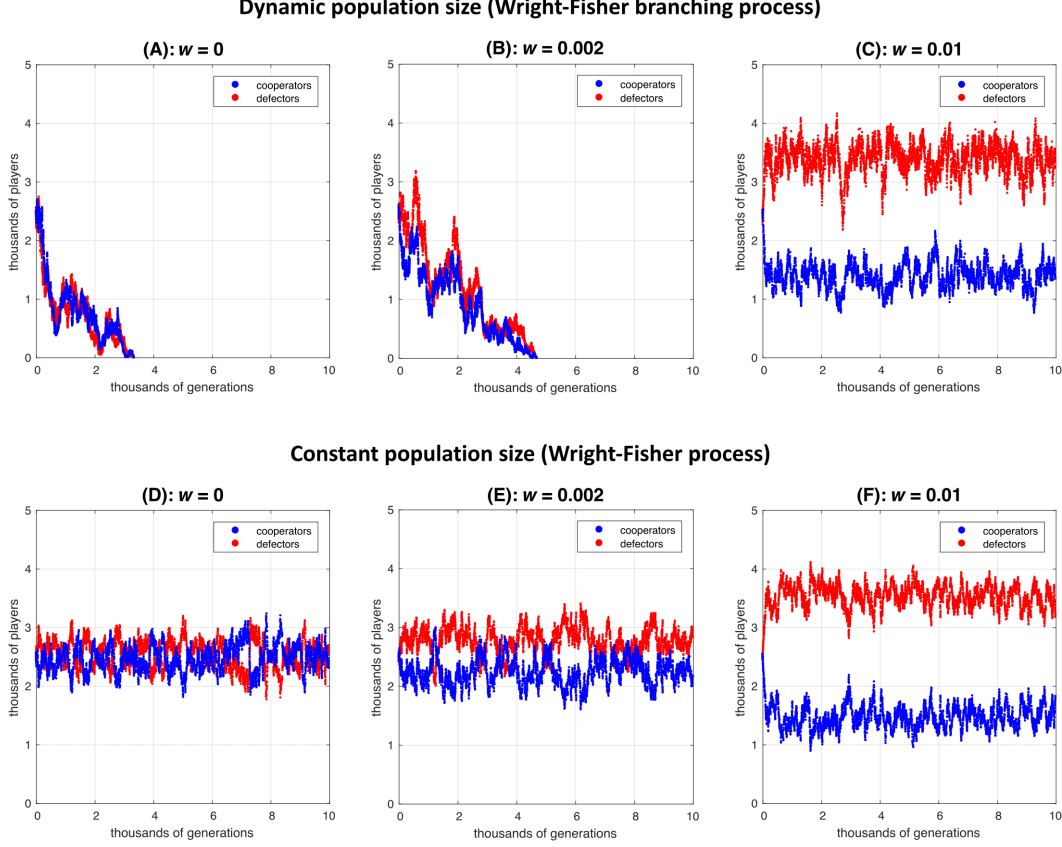**Constant population size (Wright-Fisher process)**

FIGURE 3. Mutation-selection dynamics for the Wright-Fisher branching process (panels (A)–(C)) and the Wright-Fisher process with constant population size (panels (D)–(F)). The baseline reproductive rate is $f_N = (1+r)\min\left\{1, \frac{K}{N}\right\}$, where $r = -0.001$ and $K = 5000$, although this rate is relevant only for (A)–(C) since the population size is held fixed in (D)–(F). In all panels, the population initially consists of 2500 cooperators and 2500 defectors, the strategy mutation rate is $u = 0.01$, and the multiplication factor for the public goods game is $R = 2.0$. In the top row, the population size can fluctuate over time; in the bottom row, it is held constant at 5000 players. In (A) and (D), there are no selective differences between cooperators and defectors ($w = 0$). In (B) and (E), a cooperator sacrifices a small fraction, $w = 0.002$, of his or her baseline reproductive rate in order to provide the community with a benefit. In (C) and (F), cooperators sacrifice a larger portion, $w = 0.01$, of their expected reproductive rate when provisioning a public good. While the population is artificially prevented from going extinct in (D)–(F), it can go extinct in (A)–(C) and does so quickly when the cost of cooperation is too small (panels (A) and (B)) since $f_1 < 1$. Although increasing the cost of cooperation tends to decrease the frequency of cooperators relative to defectors, a sufficient amount of selection is necessary for the survival of the population as a whole. Therefore, there is an optimal cost of cooperation, w^*, which in this case falls between 0.002 and 0.01, that maximizes the frequency of cooperators subject to survival of the population.

The dynamics of this public goods game result from the balance among three factors: mutation, selection, and population survival. Although population survival can be achieved by increasing the cost of cooperation, w, it can be destroyed by decreasing the mutation rate, u (see **Fig. SI.3**(A)). In **Supporting Information**, we show that for any $N \geqslant 1$ and any non-zero mutation rate and cost of cooperation, there exists a critical multiplication factor, R_N^*, such that the population is supported at a metastable equilibrium consisting of at
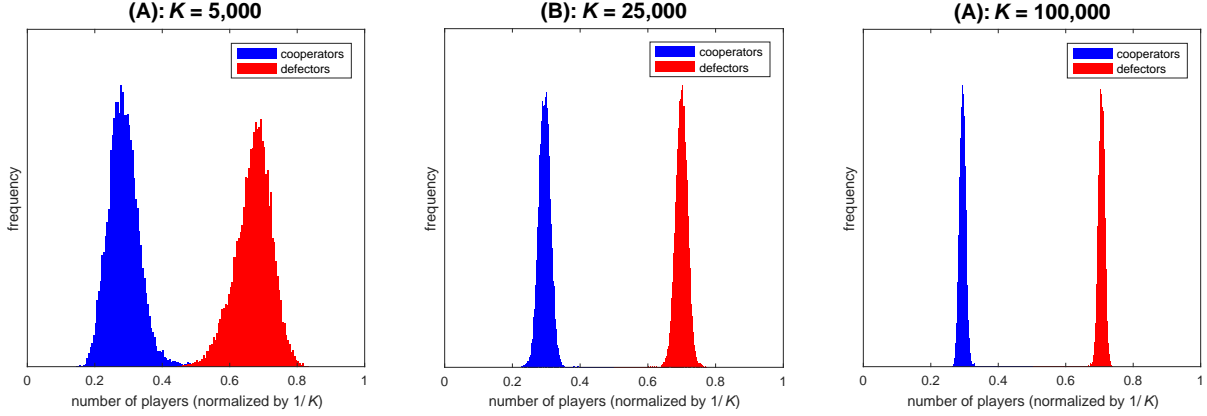
FIGURE 4. Simulation of the quasi-stationary distribution for several values of K. Here, K enters in the baseline reproductive rate, $f_N = (1 + r) \min \left\{ 1, \frac{K}{N} \right\}$, where $r = -0.001$. In each panel, cooperators and defectors are each initialized at an equal abundance of $K/2$. The plots are histograms for cooperator (blue) and defector (red) abundance over the first 25000 generations. The equilibrium fraction of cooperators, p, depends on only u, w, and R and is the same for all panels. Therefore, the peaks are centered at $x_C^*/K = 0.999x \left(1 + w \left(R - 1 \right) x \right) \approx 0.2954$ for cooperators and $x_D^*/K = 0.999 \left(1 - x \right) \left(1 + w \left(R - 1 \right) x \right) \approx 0.7066$ for defectors (see **Supporting Information**). As K grows, this quasi-stationary distribution converges to the Dirac measure centered on $(0.2954, 0.7066)$. Parameters: $u = w = 0.01$ and $R = 2.0$.

least N players whenever $R > R_N^*$. **Fig. SI.3** illustrates the effects of increasing the multiplication factor of the public good on survival. In general, the harmful effects (e.g. population extinction) of either low costs of cooperation or low mutation rates can be mitigated by increasing the return on investment in the public good, represented here by the multiplication factor, R.

When a population of defectors does not immediately go extinct, selection can increase cooperator abundance while decreasing their relative frequency (**Fig. 5**). This phenomenon is a consequence of the fact that the presence of cooperators can change the carrying capacity of population, for example so that it becomes greater than K. The fact that abundance and frequency can move in opposite directions is unique to models with variable population size and presents an interesting question about the definition of cooperator success. We show in **Supporting Information** that the fraction of cooperators present at a metastable equilibrium is independent of f_N and depends on just u, w, and R. Thus, when u, w, and R are fixed, defectors claim a fixed fraction (of at least $1/2$) of the total population, which means that cooperators are disfavored relative to defectors. However, from the perspective of population growth alone, cooperators could be considered to be favored by selection in an absolute sense because their abundance is an increasing function of the cost of cooperation.

4. DISCUSSION

In populations of fixed size, extinction is impossible and defectors can survive without the support of cooperators. This point marks perhaps the most prominent feature of classical models in evolutionary game theory that breaks down when the population size can fluctuate over time: when defectors cannot sustain themselves on their own, cooperators must be present and selection must be sufficiently strong in order to maintain the existence of the population (**Fig. 3**). Furthermore, when the size is fixed, selection decreases the frequency of cooperators if and only if it decreases the quantity of cooperators. In fluctuating populations, selection can decrease the frequency of cooperators while increasing their abundance (**Fig. 5**).

We have described w as the "cost of cooperation" instead of "selection strength" because of w's interpretation as the expected fraction of surviving offspring that must be sacrificed in order to cooperate. Since this fraction of an individual's baseline reproductive rate is shared within the population, a higher fraction means more cooperation. Defection, on the other hand, corresponds to $w = 0$ (no cooperation). Since f_N is irrelevant when one conditions on constant population size, in the standard Wright-Fisher process we can
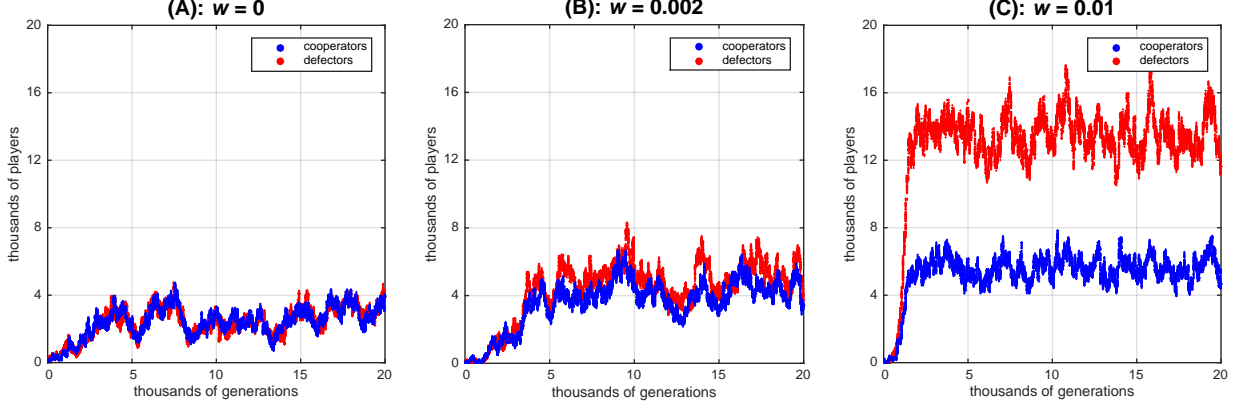
FIGURE 5. Mutation-selection dynamics when a population of defectors can sustain itself at a positive carrying capacity ($f_1 \geqslant 1$). Here, the baseline reproductive rate is $f_N = \max\left\{0, f_K + r\left(1 - \frac{N}{K}\right)\right\}$, where $f_K = 1$, $r = 0.001$, and $K = 5000$. In each panel, the population starts out with 100 cooperators and 100 defectors. In (A), $w = 0$ and cooperators are indistinguishable from defectors. The population grows to a metastable state with roughly equal frequencies of cooperators and defectors; $(x_C^*, x_D^*) = (2500, 2500)$. In (B), the cost of cooperation is positive ($w = 0.002$) and defectors begin to outnumber cooperators; $(x_C^*, x_D^*) \approx (4287, 5219)$. At the metastable state, however, cooperators in (B) slightly outnumber cooperators in (A). In (C), the cost of cooperation is further increased to $w = 0.01$ and gap between the abundance of cooperators and defectors widens; $(x_C^*, x_D^*) \approx (5806, 13890)$. Although cooperators are less frequent than defectors in (C), they are more abundant in (C) than they were in (A) and (B), which suggests that their abundance is favored by selection even though their relative frequency is not. Parameters: $u = 0.01$ and $R = 2.0$.

write the relative reproductive rates of cooperators and defectors as $F_C = 1 + w\pi_C$ and $F_D = 1 + w\pi_D$, respectively. In this context, w is typically referred to as "selection strength" [49, 56, 57] or "selection intensity" [7, 58, 59]. Therefore, the cost of cooperation–which we believe is a natural description of w for the Wright-Fisher branching process–specializes to the well-known notion of selection intensity when the population size is held constant.

That the frequencies of cooperators relative to defectors in the metastable equilibria of **Figs. 3**(C), **3**(F), and **5**(C) are all the same is not a coincidence. In fact, the fraction of cooperators, p, present at a metastable state is independent of the baseline reproductive rate, f_N, and depends on only the mutation rate, u; the cost of cooperation, w; and the multiplication factor for the public goods game, R. The population size at a metastable state is determined by f_N. We give an explicit formula for p and a condition for the existence of a non-zero metastable equilibrium in terms of f_N, w, R, and p in **Supporting Information**.

In the absence of mutation, either cooperators or defectors must be extinct in any metastable equilibrium. Although defectors outperform cooperators in a mixed population, a population of only cooperators reaches a higher carrying capacity and persists at this capacity for a longer time than does a population of defectors. Small populations of cooperators have a distinct advantage over their all-defector counterparts due to larger (average) growth rates. In particular, quick extinction is less likely for all-cooperator populations than it is for all-defector populations, which reflects an observation made by Huang et al. [22] for a related model.

Since population size can fluctuate in our model, one could also allow the multiplication factor of the public good, represented here by R, to change with N. If this multiplication factor gets weaker as N grows, then one observes dynamics similar to those here even if f_N is independent of N. Viewing R as a function of N presents an alternative way to model populations that cannot have unbounded growth due to environmental constraints. Regardless of the way in which it is framed (f_N or R depending on N), this model highlights several important results that become apparent only when the population size can change from one generation to the next.

Incorporating dynamic population size into classical evolutionary models nuances the analysis of their evolutionary dynamics. Notably, how one measures the evolutionary success of cooperators is not as unambiguous here as it is for fixed-size models. Selection can favor cooperator abundance despite disfavoring cooperator frequency. Cooperators are exploited by defectors, yet they can be crucial to the maintenance of the population as a whole. These observations mark important baseline characteristics of well-mixed populations that are useful for studying how other features, such as population structure, further affect the evolution of cooperation. While (even idealized) evolutionary models with fluctuating population size undoubtedly present technical obstacles not apparent in their static counterparts, they more accurately represent natural populations and form an important area for further development in evolutionary game theory.

References

[1] R. Axelrod and W. Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, Mar 1981. doi: 10.1126/science.7466396.

[2] M. A. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Belknap Press, 2006.

[3] M. A. Nowak. Five rules for the evolution of cooperation. *Science*, 314(5805):1560–1563, Dec 2006. doi: 10.1126/science.1133755.

[4] P. D. Taylor and L. B. Jonker. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1-2):145–156, jul 1978. doi: 10.1016/0025-5564(78)90077-9.

[5] C. Hauert, M. Holmes, and M. Doebeli. Evolutionary games and population dynamics: maintenance of cooperation in public goods games. *Proceedings of the Royal Society B: Biological Sciences*, 273(1605): 3131–3132, Dec 2006. doi: 10.1098/rspb.2006.3717.

[6] P. A. P. Moran. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54(01):60, Jan 1958. doi: 10.1017/s0305004100033193.

[7] M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428(6983):646–650, Apr 2004. doi: 10.1038/nature02414.

[8] C. Taylor, D. Fudenberg, A. Sasaki, and M. A. Nowak. Evolutionary game dynamics in finite populations. *Bulletin of Mathematical Biology*, 66(6):1621–1644, Nov 2004. doi: 10.1016/j.bulm.2004.03.004.

[9] E. Lieberman, C. Hauert, and M. A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433(7023): 312–316, Jan 2005. doi: 10.1038/nature03204.

[10] H. Ohtsuki, C. Hauert, E. Lieberman, and M. A. Nowak. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441(7092):502–505, May 2006. doi: 10.1038/nature04605.

[11] P. D. Taylor, T. Day, and G. Wild. Evolution of cooperation in a finite homogeneous graph. *Nature*, 447(7143):469–472, May 2007. doi: 10.1038/nature05784.

[12] G. Szabó and G. Fáth. Evolutionary games on graphs. *Physics Reports*, 446(4-6):97–216, Jul 2007. doi: 10.1016/j.physrep.2007.04.004.

[13] C. E. Tarnita, T. Antal, H. Ohtsuki, and M. A. Nowak. Evolutionary dynamics in set structured populations. *Proceedings of the National Academy of Sciences*, 106(21):8601–8604, May 2009. doi: 10.1073/pnas.0903019106.

[14] M. A. Nowak, C. E. Tarnita, and T. Antal. Evolutionary dynamics in structured populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537):19–30, Nov 2009. doi: 10.1098/rstb.2009.0215.

[15] C. Hauert and L. Imhof. Evolutionary games in deme structured, finite populations. *Journal of Theoretical Biology*, 299:106–112, Apr 2012. doi: 10.1016/j.jtbi.2011.06.010.

[16] F. Débarre, C. Hauert, and M. Doebeli. Social evolution in structured populations. *Nature Communications*, 5, Mar 2014. doi: 10.1038/ncomms4409.

[17] M. Kimmel and D. E. Axelrod. *Branching Processes in Biology.* Springer New York, 2015. doi: 10.1007/978-1-4939-1559-0.

[18] A. Melbinger, J. Cremer, and E. Frey. Evolutionary game theory in growing populations. *Physical Review Letters*, 105(17), Oct 2010. doi: 10.1103/physrevlett.105.178101.

[19] S. Novak, K. Chatterjee, and M. A. Nowak. Density games. *Journal of Theoretical Biology*, 334:26–34, Oct 2013. doi: 10.1016/j.jtbi.2013.05.029.

[20] G. W. A. Constable, T. Rogers, A. J. McKane, and C. E. Tarnita. Demographic noise can reverse the direction of deterministic selection. *Proceedings of the National Academy of Sciences*, 113(32): E4745–E4754, Jul 2016. doi: 10.1073/pnas.1603693113.

[21] P. Ashcroft, C. E. R. Smith, M. Garrod, and T. Galla. Effects of population growth on the success of invading mutants. *Journal of Theoretical Biology*, 420:232–240, May 2017. doi: 10.1016/j.jtbi.2017.03.014.

[22] W. Huang, C. Hauert, and A. Traulsen. Stochastic game dynamics under demographic fluctuations. *Proceedings of the National Academy of Sciences*, 112(29):9064–9069, Jul 2015. doi: 10.1073/pnas.1418745112.

[23] P. Czuppon and A. Traulsen. Fixation probabilities in populations under demographic fluctuations. arXiv preprint arXiv:1708.09665, 2017.

[24] R. A. Fisher. *The Genetical Theory of Natural Selection.* Clarendon Press, 1930.

[25] S. Wright. Evolution in Mendelian populations. *Genetics*, 16:97–159, 1931.

[26] W. J. Ewens. *Mathematical Population Genetics.* Springer New York, 2004. doi: 10.1007/978-0-387-21822-9.

[27] L. A. Imhof and M. A. Nowak. Evolutionary game dynamics in a Wright-Fisher process. *Journal of Mathematical Biology*, 52(5):667–681, Feb 2006. doi: 10.1007/s00285-005-0369-8.

[28] O. P. Hauser, A. Traulsen, and M. A. Nowak. Heterogeneity in background fitness acts as a suppressor of selection. *Journal of Theoretical Biology*, 343:178–185, Feb 2014. doi: 10.1016/j.jtbi.2013.10.013.

[29] Y. Iwasa and F. Michor. Evolutionary Dynamics of Intratumor Heterogeneity. *PLoS ONE*, 6(3):e17866, Mar 2011. doi: 10.1371/journal.pone.0017866.

[30] K. Sigmund. *The calculus of selfishness.* Princeton University Press, 2010.

[31] R. Craig MacLean and C. Brandon. Stable public goods cooperation and dynamic social interactions in yeast. *Journal of Evolutionary Biology*, 21(6):1836–1843, Nov 2008. doi: 10.1111/j.1420-9101.2008.01579.x.

[32] T. Czárán and R. F. Hoekstra. Microbial Communication, Cooperation and Cheating: Quorum Sensing Drives the Evolution of Cooperation in Bacteria. *PLoS ONE*, 4(8):e6655, Aug 2009. doi: 10.1371/journal.pone.0006655.

[33] O. X. Cordero, L.-A. Ventouras, E. F. DeLong, and M. F. Polz. Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proceedings of the National Academy of Sciences*, 109(49):20059–20064, Nov 2012. doi: 10.1073/pnas.1213344109.

[34] A. Sanchez and J. Gore. Feedback between Population and Evolutionary Dynamics Determines the Fate of Social Microbial Populations. *PLoS Biology*, 11(4):e1001547, Apr 2013. doi: 10.1371/journal.pbio.1001547.

[35] B. Allen, J. Gore, and M. A. Nowak. Spatial dilemmas of diffusible public goods. *eLife*, 2, Dec 2013. doi: 10.7554/elife.01169.

[36] R. Seyfarth, D. Cheney, and P. Marler. Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science*, 210(4471):801–803, Nov 1980. doi: 10.1126/science.7433999.

[37] T. Clutton-Brock. *Mammal Societies.* Wiley, 2016.

[38] T. H. Clutton-Brock, M. J. O'Riain, P. N. M. Brotherton, D. Gaynor, R. Kansky, A. S. Griffin, and M. Manser. Selfish Sentinels in Cooperative Mammals. *Science*, 284(5420):1640–1644, Jun 1999. doi: 10.1126/science.284.5420.1640.

[39] M. Milinski, D. Semmann, H.-J. Krambeck, and J. Marotzke. Stabilizing the earths climate is not a losing game: Supporting evidence from public goods experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 103(11):3994–3998, 2006. doi: 10.1073/pnas.0504902103.

[40] J. Jacquet, K. Hagel, C. Hauert, J. Marotzke, T. Röhl, and M. Milinski. Intra- and intergenerational discounting in the climate game. *Nature Climate Change*, 3(12):1025–1028, Oct 2013. doi: 10.1038/nclimate2024.

[41] M. D. Ehmke and J. F. Shogren. Experimental methods for environment and development economics. *Environment and Development Economics*, 14(04):419, Aug 2008. doi: 10.1017/s1355770x08004592.

[42] M. Archetti and I. Scheuring. Review: Game theory of public goods in one-shot social dilemmas without assortment. *Journal of Theoretical Biology*, 299:9–20, Apr 2012. doi: 10.1016/j.jtbi.2011.06.018.

[43] C. Hauert and G. Szabó. Prisoners dilemma and public goods games in different geometries: Compulsory versus voluntary interactions. *Complexity*, 8:31–38, 2003. doi: 10.1002/cplx.10092.

[44] M. Doebeli, Y. Ispolatov, and B. Simon. Towards a mechanistic foundation of evolutionary theory. *eLife*, 6, Feb 2017. doi: 10.7554/elife.23804.

[45] P. Haccou, P. Jagers, and V. A. Vatutin. *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press, 2005. doi: 10.1017/CBO9780511629136.

[46] X. Chen, Y. Liu, Y. Zhou, L. Wang, and M. Perc. Adaptive and Bounded Investment Returns Promote Cooperation in Spatial Public Goods Games. *PLoS ONE*, 7(5):e36895, May 2012. doi: 10.1371/journal.pone.0036895.

[47] M. A. Nowak, S. Bonhoeffer, and R. M. May. More spatial games. *International Journal of Bifurcation and Chaos*, 04(01):33–56, Feb 1994. doi: 10.1142/s0218127494000046.

[48] C. E. Tarnita, H. Ohtsuki, T. Antal, F. Fu, and M. A. Nowak. Strategy selection in structured populations. *Journal of Theoretical Biology*, 259(3):570–581, Aug 2009. doi: 10.1016/j.jtbi.2009.03.035.

[49] T. Antal, A. Traulsen, H. Ohtsuki, C. E. Tarnita, and M. A. Nowak. Mutation-selection equilibrium in games with multiple strategies. *Journal of Theoretical Biology*, 258(4):614–622, Jun 2009. doi: 10.1016/j.jtbi.2009.02.010.

[50] C. E. Tarnita, N. Wage, and M. A. Nowak. Multiple strategies in structured populations. *Proceedings of the National Academy of Sciences*, 108(6):2334–2337, Jan 2011. doi: 10.1073/pnas.1016008108.

[51] B. Wu, C. S. Gokhale, L. Wang, and A. Traulsen. How small are small mutation rates? *Journal of Mathematical Biology*, 64(5):803–827, May 2011. doi: 10.1007/s00285-011-0430-8.

[52] A. Traulsen, C. Hauert, H. De Silva, M. A. Nowak, and K. Sigmund. Exploration dynamics in evolutionary games. *Proceedings of the National Academy of Sciences*, 106(3):709–712, Jan 2009. doi: 10.1073/pnas.0808450106.

[53] P. Jagers and F. C. Klebaner. Dependence and Interaction in Branching Processes. In *Springer Proceedings in Mathematics & Statistics*, pages 325–333. Springer Science + Business Media, Nov 2012. doi: 10.1007/978-3-642-33549-5_19.

[54] K. Hamza, P. Jagers, and F. C. Klebaner. On the establishment, persistence, and inevitable extinction of populations. *Journal of Mathematical Biology*, 72(4):797–820, Jun 2015. doi: 10.1007/s00285-015-0903-2.

[55] P. Jagers and F. C. Klebaner. Population-size-dependent, age-structured branching processes linger around their carrying capacity. *Journal of Applied Probability*, 48A(0):249–260, Aug 2011. doi: 10.1239/jap/1318940469.

[56] E. Chastain, A. Livnat, C. Papadimitriou, and U. Vazirani. Algorithms, games, and evolution. *Proceedings of the National Academy of Sciences*, 111(29):10620–10623, Jun 2014. doi: 10.1073/pnas.1406556111.

[57] B. Allen, G. Lippner, Y.-T. Chen, B. Fotouhi, N. Momeni, S.-T. Yau, and M. A. Nowak. Evolutionary dynamics on any population structure. *Nature*, 544(7649):227–230, Mar 2017. doi: 10.1038/nature21723.

[58] B. Wu, P. M. Altrock, L. Wang, and A. Traulsen. Universality of weak selection. *Physical Review E*, 82(4), Oct 2010. doi: 10.1103/physreve.82.046106.

[59] B. Wu, J. García, C. Hauert, and A. Traulsen. Extrapolating weak selection in evolutionary games. *PLoS Computational Biology*, 9(12):e1003381, Dec 2013. doi: 10.1371/journal.pcbi.1003381.

[60] T. Antal, H. Ohtsuki, J. Wakeley, P. D. Taylor, and M. A. Nowak. Evolution of cooperation by phenotypic similarity. *Proceedings of the National Academy of Sciences*, 106(21):8597–8600, Apr 2009. doi: 10.1073/pnas.0902528106.

[61] C. Veller and L. K. Hayward. Finite-population evolution with rare mutations in asymmetric games. *Journal of Economic Theory*, 162:93–113, Mar 2016. doi: 10.1016/j.jet.2015.12.005.

SI.1. WRIGHT-FISHER BRANCHING PROCESS

SI.1.1. **Update rule.** Suppose that the population is well-mixed but allowed to vary in size. For simplicity, assume that we are dealing with a symmetric game with two strategies, C ("cooperate") and D ("defect"). A state of the population is then uniquely defined by a pair, (x_C, x_D), where x_C and x_C are the number of players using C and D, respectively. Thus, the population size is $x_C + x_D$, which can vary from generation to generation.

Suppose that the reproductive rates of cooperators and defectors in state (x_C, x_D) are given by functions $F_C = F_C(x_C, x_D)$ and $F_D = F_D(x_C, x_D)$, respectively. That is, the reproductive rates are frequency-dependent and determined by the number of each type of player in the population. Moreover, suppose that we wish to interpret each reproductive rate as the expected number of surviving offspring of an individual over its lifetime, i.e. the "absolute" interpretation of reproductive rate. We assume a reproductive mechanism in which the number of offspring is Poisson-distributed at the parent's reproductive rate. Therefore, the probability of transitioning from state (x_C, x_D) to state (y_C, y_D) is

$$\mathbf{P}_{\mathrm{WFB}}\Big((x_C, x_D), (y_C, y_D) \Big) = \left(\frac{(x_C F_C)^{y_C} e^{-x_C F_C}}{y_C!} \right) \left(\frac{(x_D F_D)^{y_D} e^{-x_D F_D}}{y_D!} \right). \tag{SI.1}$$

Note that this update rule for the "Wright-Fisher branching process" defines transitions over generations under the assumption that these generations are non-overlapping.

Remark 1. If the population size is static and fixed at N, then, for $x_C + x_D = y_C + y_D = N$,

$$\mathbf{P}_{\mathrm{WFB}}\Big((x_C, x_D), (y_C, y_D) \mid \text{population size is static} \Big)$$

$$= \frac{\left(\frac{(x_C F_C)^{y_C} e^{-x_C F_C}}{y_C!} \right) \left(\frac{(x_D F_D)^{y_D} e^{-x_D F_D}}{y_D!} \right)}{\sum_{z_C + z_D = N} \left(\frac{(x_C F_C)^{z_C} e^{-x_C F_C}}{z_C!} \right) \left(\frac{(x_D F_D)^{z_D} e^{-x_D F_D}}{z_D!} \right)}$$

$$= \frac{\left(\frac{(x_C F_C)^{y_C} e^{-x_C F_C}}{y_C!} \right) \left(\frac{(x_D F_D)^{y_D} e^{-x_D F_D}}{y_D!} \right)}{\frac{1}{N!} \sum_{z_C + z_D = N} N! \left(\frac{(x_C F_C)^{z_C} e^{-x_C F_C}}{z_C!} \right) \left(\frac{(x_D F_D)^{z_D} e^{-x_D F_D}}{z_D!} \right)}$$

$$= \frac{\left(\frac{(x_C F_C)^{y_C} e^{-x_C F_C}}{y_C!} \right) \left(\frac{(x_D F_D)^{y_D} e^{-x_D F_D}}{y_D!} \right)}{\frac{1}{N!} (x_C F_C + x_D F_D)^N e^{-(x_C F_C + x_D F_D)}}$$

$$= \frac{N!}{y_C! y_D!} \left(\frac{x_C F_C}{x_C F_C + x_D F_D} \right)^{y_C} \left(\frac{x_D F_D}{x_C F_C + x_D F_D} \right)^{y_D}$$

$$= \mathbf{P}_{\mathrm{WF}}(x_C, y_C), \tag{SI.2}$$

recovering the classical transition rule based on binomial sampling. Therefore, the update rule defined by **Eq. SI.1** produces a process that generalizes the classical, frequency-dependent Wright-Fisher process.

SI.1.2. **Reproductive rates and selection.** In the neutral model, each player in a population of size N has a reproductive rate determined by a baseline reproductive rate, f_N. We give two examples of f_N below.

SI.1.2.1. *Rectified linear rate.* One natural way to model reproductive rate is as a linear function of the population size, N. In this case, we can write $f_N = \max\left\{0, f_K + r\left(1 - \frac{N}{K}\right)\right\}$ for some parameters f_K, r, and K. We refer to f_N as a "rectified" linear rate since it piecewise-linear with the constraint $f_N \geqslant 0$ for every N. Note that $f_{N*} = 1$ when $N^* = K\left(1 - \frac{1}{r}(1 - f_K)\right)$. Therefore, N^* may be interpreted as the (neutral) carrying capacity of the population since each individual is replaced by one offspring on average. Note that K itself is not necessarily the neutral carrying capacity for this type of reproductive rate.

SI.1.2.2. *Threshold-constant rate.* If the reproductive rate is constant, then every player expects to produce $1+r$ offspring that survive into the next generation, where $r \geqslant -1$. We assume that this growth is eventually bounded by environmental constraints, so we set $f_N = (1+r) \min\{1, K/N\}$ for some K. We refer to f_N as a "threshold-constant" rate since it is constant up to a threshold ($N = K$) and then decreasing to 0 beyond K. When $r < 0$, there is no solution to $f_N = 1$ since $f_N \leqslant 1 + r < 1$ for each N. When $r > 0$, we have $f_{(1+r)K} = 1$, so $N^* = (1+r)K$ is the neutral carrying capacity.

Examples of each of these rate functions are shown in **Fig. 1** in the main text.

SI.1.2.3. *Selection.* In a game with strategies C and D, let $\pi_C(x_C, y_C)$ and $\pi_D(x_C, x_D)$ be the total payoffs to C and D, respectively, when there are x_C cooperators and x_D defectors. If the population size is fixed, then payoff, π, is commonly converted to fitness , f, by defining $f = 1 + w\pi$, where w is the so-called "selection strength" [see 49, 60]. This perturbation has even been extended to asymmetric games played between different populations [61]. While our setup is somewhat different, we maintain this convention of using payoffs from a game to perturb reproductive rates. In particular, if w is a parameter representing the intensity of selection, then the reproductive rates of cooperators and defectors are given by

$$F_C(x_C, x_D) := \Big(1 + w\pi_C(x_C, x_D)\Big)f_{x_C + x_D}; \tag{SI.3a}$$

$$F_D(x_C, x_D) := \Big(1 + w\pi_D(x_C, x_D)\Big)f_{x_C + x_D}, \tag{SI.3b}$$

respectively. In other words, the baseline reproductive rate, $f_{x_C + x_D}$, is perturbed by the game according to the strength of selection, w. In order to maintain non-negative reproductive rates, w must be sufficiently small. In the next section, we consider a public goods game in which w has a clear biological interpretation.

SI.2. Dynamics of the public goods game

In the public goods game, a cooperator sacrifices a fraction, w, of his or her baseline reproductive rate in order to provision a public good. A cooperator's contribution is enhanced by a factor of $R > 1$ (the multiplication factor of the public good), which is then distributed evenly among all of the players in the population. In terms of the payoff function in **Eq. SI.3a**, we have $\pi_C(x_C, x_D) = R\left(\frac{x_C}{x_C + x_D}\right) - 1$ and $\pi_D(x_C, x_D) = R\left(\frac{x_C}{x_C + x_D}\right)$ as well as **Eq. 4** in the main text.

SI.2.1. **Metastable equilibria.** Consider a population evolving according to **Eq. SI.1**. As noted in the main text, such a branching process either grows without bound or goes extinct. Even when the population has an extinction probability of 1, there can be so-called "metastable" states (or "equilibria") around which the population fluctuates for many generations. While a quasi-stationary distribution for the process describes the distribution of strategy abundances prior to extinction, a metastable state describes the mean(s) around which these strategy counts fluctuate. Here, we are interested in understanding when these metastable states exist and how they are influenced by the parameters of the model.

Let $\mathbf{E}_{(x_C, x_D)}[y_C]$ (resp. $\mathbf{E}_{(x_C, x_D)}[y_D]$) denote the expected abundance of cooperators (resp. defectors) in the next generation given x_C cooperators and x_D defectors in the current generation. Formally, a metastable equilibrium for this process is a state at which $\mathbf{E}_{(x_C^*, x_D^*)}[y_C] = x_C^*$ and $\mathbf{E}_{(x_C^*, x_D^*)}[y_D] = x_D^*$. In other words, the abundances of cooperators and defectors each remain unchanged (on average) at a metastable state. The term "metastable" is used because the population both fluctuates around this state and eventually goes extinct. We discuss the extinction time later in §SI.3. First, we derive the metastable equilibria for public goods games.

SI.2.1.1. *Derivation of metastable equilibria.* Let u be the strategy-mutation rate. With probability $1-u$, an offspring acquires the strategy of the parent. With probability u, the offspring takes on one of C and D uniformly at random. In state (x_C, x_D), the expected number of cooperators in the next generation is

$$\mathbf{E}_{(x_C, x_D)}[y_C] = \sum_{(y_C, y_D)} \mathbf{P}_{\mathrm{WFB}}\Big((x_C, x_D), (y_C, y_D)\Big)\left(\left(1 - \frac{u}{2}\right)y_C + \left(\frac{u}{2}\right)y_D\right)$$

$$= \sum_{(y_C, y_D)} \left(\frac{(x_C F_C)^{y_C}\, e^{-x_C F_C}}{y_C!}\right)\left(\frac{(x_D F_D)^{y_D}\, e^{-x_D F_D}}{y_D!}\right)\left(\left(1 - \frac{u}{2}\right)y_C + \left(\frac{u}{2}\right)y_D\right)$$

$$= \left(1 - \frac{u}{2}\right) x_C F_C + \left(\frac{u}{2}\right) x_D F_D. \tag{SI.4}$$

Similarly, the expected number of defectors in the next generation is $\left(\frac{u}{2}\right) x_C F_C + \left(1 - \frac{u}{2}\right) x_D F_D$. Therefore, the system of equations we need to solve in order to find a metastable equilibrium is

$$x_C = \left(1 - \frac{u}{2}\right) x_C F_C + \left(\frac{u}{2}\right) x_D F_D; \tag{SI.5a}$$

$$x_D = \left(\frac{u}{2}\right) x_C F_C + \left(1 - \frac{u}{2}\right) x_D F_D. \tag{SI.5b}$$

In other words, it must be true that

$$x_C = \left[\left(1 - \frac{u}{2}\right) x_C \Big(1 + w\pi_C\left(x_C, x_D\right)\Big) + \left(\frac{u}{2}\right) x_D \Big(1 + w\pi_D\left(x_C, x_D\right)\Big)\right] f_{x_C+x_D}; \tag{SI.6a}$$

$$x_D = \left[\left(\frac{u}{2}\right) x_C \Big(1 + w\pi_C\left(x_C, x_D\right)\Big) + \left(1 - \frac{u}{2}\right) x_D \Big(1 + w\pi_D\left(x_C, x_D\right)\Big)\right] f_{x_C+x_D}. \tag{SI.6b}$$

These equations are always satisfied when $x_C = x_D = 0$ (population extinction). There can also be solutions to **Eq. SI.6** with $x_C \neq 0$ or $x_D \neq 0$; we give a condition for the existence of non-zero solutions below.

Remark 2. If $u = 0$, then **Eq. SI.6** reduces to the system

$$x_C = \Big(1 + w\pi_C\left(x_C, x_D\right)\Big) x_C f_{x_C+x_D}; \tag{SI.7a}$$

$$x_D = \Big(1 + w\pi_D\left(x_C, x_D\right)\Big) x_D f_{x_C+x_D}. \tag{SI.7b}$$

If x_C and x_D satisfy this system and $x_C, x_D \neq 0$, then

$$\Big(1 + w\pi_C\left(x_C, x_D\right)\Big) f_{x_C+x_D} = \Big(1 + w\pi_D\left(x_C, x_D\right)\Big) f_{x_C+x_D} = 1. \tag{SI.8}$$

Therefore, either $w = 0$ and $f_{x_C+x_D} = 1$ or $w \neq 0$ and $\pi_C\left(x_C, x_D\right) = \pi_D\left(x_C, x_D\right)$. However, for the public goods game, it is always the case that $\pi_C\left(x_C, x_D\right) < \pi_D\left(x_C, x_D\right)$ when $x_C > 0$, so it must be true that $w = 0$ and $f_{x_C+x_D} = 1$. Thus, if $u = 0$ and $w > 0$, then any solution satisfies $x_C = 0$ or $x_D = 0$. In other words, in the absence of mutation, selection forces the extinction of at least one strategy.

Similarly, if $u > 0$ and there is a solution (x_C, x_D) with $x_C = 0$, then one obtains $x_D F_D = 0$ from **Eq. SI.5a** and $x_D = 0$ from **Eq. SI.5b**. For a similar reason, if there is solution with $x_D = 0$, then it must also be true that $x_C = 0$. Thus, if (x_C, x_D) is a non-zero solution to **Eq. SI.6**, then $x_C > 0$ and $x_D > 0$.

Lemma 1. If $u > 0$, then there exists $p \in (0, 1)$ such that any non-zero solution to **Eq. SI.6**, (x_C, x_D), satisfies $\frac{x_C}{x_C+x_D} = p$. In other words, the fraction of cooperators is the same in any solution to **Eq. SI.6**.

Proof. If (x_C, x_D) is a solution to **Eq. SI.6** with $x_C, x_D \geqslant 0$ and $x_C + x_D > 0$, then, with $p := \frac{x_C}{x_C+x_D}$,

$$
\begin{aligned}
p &= \frac{x_C}{x_C + x_D} \\
&= \frac{\left(1 - \frac{u}{2}\right) x_C \Big(1 + w\pi_C\left(x_C, x_D\right)\Big) + \left(\frac{u}{2}\right) x_D \Big(1 + w\pi_D\left(x_C, x_D\right)\Big)}{x_C \Big(1 + w\pi_C\left(x_C, x_D\right)\Big) + x_D \Big(1 + w\pi_D\left(x_C, x_D\right)\Big)} \\
&= \frac{\left(1 - \frac{u}{2}\right) \left(\frac{x_C}{x_C+x_D}\right) \Big(1 + w\pi_C\left(\frac{x_C}{x_C+x_D}, \frac{x_D}{x_C+x_D}\right)\Big) + \left(\frac{u}{2}\right) \left(\frac{x_D}{x_C+x_D}\right) \Big(1 + w\pi_D\left(\frac{x_C}{x_C+x_D}, \frac{x_D}{x_C+x_D}\right)\Big)}{\left(\frac{x_C}{x_C+x_D}\right) \Big(1 + w\pi_C\left(\frac{x_C}{x_C+x_D}, \frac{x_D}{x_C+x_D}\right)\Big) + \left(\frac{x_D}{x_C+x_D}\right) \Big(1 + w\pi_D\left(\frac{x_C}{x_C+x_D}, \frac{x_D}{x_C+x_D}\right)\Big)} \\
&= \frac{\left(1 - \frac{u}{2}\right) p \Big(1 + w\pi_C\left(p, 1-p\right)\Big) + \left(\frac{u}{2}\right) (1-p) \Big(1 + w\pi_D\left(p, 1-p\right)\Big)}{p \Big(1 + w\pi_C\left(p, 1-p\right)\Big) + (1-p) \Big(1 + w\pi_D\left(p, 1-p\right)\Big)}.
\end{aligned}
\tag{SI.9}
$$

Since $\pi_C\left(x_C, x_D\right) = R\left(\frac{x_C}{x_C+x_D}\right) - 1$ and $\pi_D\left(x_C, x_D\right) = R\left(\frac{x_C}{x_C+x_D}\right)$, **Eq. SI.9** is equivalent to

$$\varphi\left(p\right) := 2\left(1 - uR\right) wp^2 + \left(uw\left(R + 1\right) - 2u - 2w\right) p + u = 0. \tag{SI.10}$$

16

Since $\varphi(p)$ is (at most) quadratic, $\varphi(0) = u$, and $\varphi(1) = -u(1 - w + Rw)$, we see that if $u > 0$, then there is a unique solution to **Eq. SI.9** that falls within $[0, 1]$, and, furthermore, this solution is in $(0, 1)$. Explicitly,

$$p = \frac{uw(R+1) - 2u - 2w + \sqrt{u^2 w^2 R^2 + 2u^2 w^2 R + 4u^2 wR - 4uw^2 R + u^2 w^2 - 4u^2 w + 4u^2 - 4uw^2 + 4w^2}}{4(uR - 1)w}$$

(SI.11)

if $R \neq 1/u$ and $w \neq 0$, and $p = u/(2u + (1-u)w)$ if $R = 1/u$ or $w = 0$, which completes the proof. □

From the proof of Lemma 1, we see that if $u = 0$, then either (i) $w = 0$ and every $p \in [0, 1]$ is a solution to **Eq. SI.9** or (ii) $w > 0$ and the only solutions to **Eq. SI.9** are $p = 0$ and $p = 1$. For any $u \in [0, 1]$, the fraction of cooperators in a non-zero metastable equilibrium is independent of the baseline reproductive rate, f_N. However, the existence of a metastable equilibrium and the size of the population at such an equilibrium both depend on the baseline reproductive rate. Suppose that $x_C = pN$ and $x_D = (1-p)N$ satisfy **Eq. SI.6**, where $p \in [0, 1]$ is a fraction of cooperators that satisfies **Eq. SI.9**. From **Eq. SI.6**,

$$p = \left[\left(1 - \frac{u}{2} \right) p \left(1 + w\pi_C(p, 1-p) \right) + \left(\frac{u}{2} \right)(1-p)\left(1 + w\pi_D(p, 1-p) \right) \right] f_N; \tag{SI.12a}$$

$$1 - p = \left[\left(\frac{u}{2} \right) p \left(1 + w\pi_C(p, 1-p) \right) + \left(1 - \frac{u}{2} \right)(1-p)\left(1 + w\pi_D(p, 1-p) \right) \right] f_N, \tag{SI.12b}$$

which, in turn, holds if and only if the total population size, N satisfies

$$f_N = \frac{1}{1 + w(R-1)p}. \tag{SI.13}$$

The right-hand-side of **Eq. SI.13** is independent of f_N, and once this quantity is calculated, it is straightforward to check for any f_N whether there exists N for which **Eq. SI.13** holds. If f_N is strictly monotonic, then there exists at most one N that satisfies this equation. For other types of baseline reproductive rates, there might be several such N that satisfy **Eq. SI.13** (resulting in several non-zero metastable equilibria).

In addition to the simulations described in the main text, **Figs. SI.1**–**SI.2** demonstrate further effects of model parameters on metastable equilibria.

Lemma 2. If $u \in (0, 1)$ and $w \neq 0$, then p is a strictly increasing function of R with $p \uparrow 1/2$ as $R \to \infty$.

Proof. Since the polynomial defined by **Eq. SI.10** satisfies $\varphi(0) = u$ and $\varphi(1/2) = -w(1-u)/2$, we see that the solution to $\varphi(p) = 0$ that falls within $[0, 1]$ is actually at most $1/2$. Moreover, we can write

$$\varphi_R(p) := \varphi(p) = 2(uR - 1)wp\left(\frac{1}{2} - p \right) - \left((1-u)w + 2u \right)p + u, \tag{SI.14}$$

where, notably, only the coefficient of $p\left(\frac{1}{2} - p \right)$ depends on R. Thus, if $R_1 < R_2$ and p_1 satisfies $\varphi_{R_1}(p_1) = 0$, then $\varphi_{R_2}(p_1) > 0$. Since $\varphi(1/2) < 0$, the unique solution to $\varphi_{R_2}(p_2) = 0$ satisfies $p_2 > p_1$, thus p is an increasing function of R. That $p \uparrow 1/2$ as $R \to \infty$ follows immediately from taking the limit of **Eq. SI.11**. □

Theorem 1. Suppose that $f_N \downarrow 0$ as $N \to \infty$. If $u \neq 0$ and $w \neq 0$, then, for each $N \geq 1$, there is a critical multiplication factor, $R_N^* \geq 1$, which is the minimum multiplication factor for which there exists a non-zero metastable equilibrium supporting a population size of at least N whenever $R \geq R_N^*$.

Proof. Since $p \uparrow 1/2$ as $R \to \infty$, we see that $\frac{1}{1+w(R-1)p} \downarrow 0$ as $R \to \infty$. Let

$$R_N^* := \inf \left\{ R \geq 1 \ : \ f_{N'} = \frac{1}{1 + w(R-1)p} \text{ for some } N' \geq N \right\}. \tag{SI.15}$$

Since $f_N \downarrow 0$ as $N \to \infty$, we have $R_N^* < \infty$. If $R \geq R_N^*$ and N' satisfies $f_{N'} = \frac{1}{1+w(R_N^*-1)p}$, then $(x_C^*, x_D^*) = (pN', (1-p)N')$ is a metastable equilibrium by **Eq. SI.12**. Furthermore, if $R \geq R_N^*$, then

$$\frac{1}{1 + w(R-1)p} \leq \frac{1}{1 + w(R_N^* - 1)p}, \tag{SI.16}$$

and it follows that any solution to $f_{N''} = \frac{1}{1+w(R-1)p}$ satisfies $N'' \geq N' \geq N$, as desired. □

Fig. SI.3 illustrates the effects of increasing R on population survival.
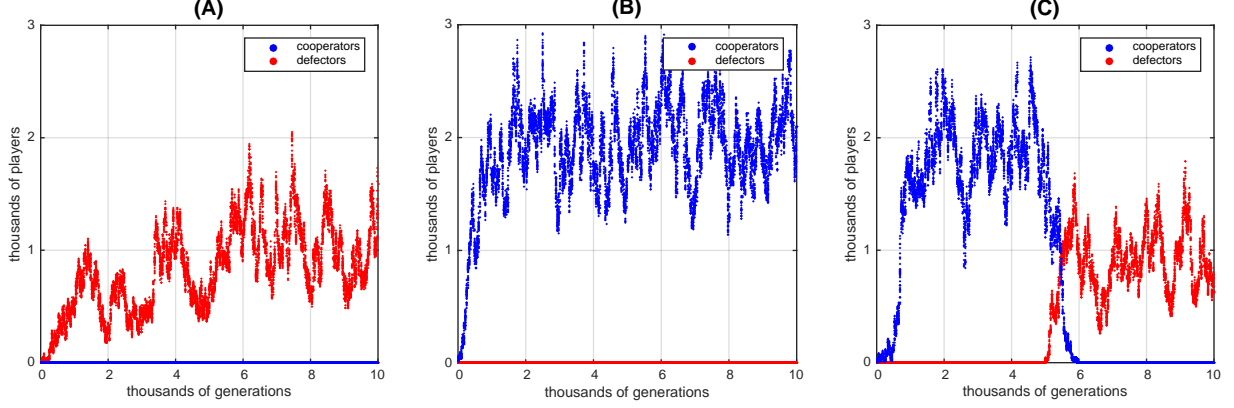
FIGURE SI.1. Evolutionary game dynamics when there is no mutation and the baseline reproductive rate, f_N, is defined by $f_N = \max\{0, f_K + r(1 - \frac{N}{K})\}$, where $f_K = 1$, $r = 0.005$, and $K = 1000$. The multiplication factor for the public goods game is $R = 1.5$ and the cost of cooperation is $w = 0.01$. In (A), the population is initialized with no cooperators and 10 defectors. The defectors grow until they reach their carrying capacity of 1000 and then persist at this metastable equilibrium ($f_1 \geqslant 1$). In (B), the population is initialized with 10 cooperators and no defectors. The cooperators then grow in abundance until they reach their carrying capacity of approximately 2000 players. It is immediate from panels (A) and (B) that groups of cooperators perform better than groups of defectors since selection allows them to reach a higher carrying capacity. In (C), the population is initialized with 10 cooperators and no defectors, and the population then proceeds to reach its carrying capacity. After 5000 generations, an additional 10 defectors are introduced into the population, which disrupts the metastable equilibrium reached by the all-cooperator population. Defectors then outcompete and replace cooperators and finally reach their carrying capacity, which, as in (A), is significantly lower than the carrying capacity of an all-cooperator population.

Remark 3. If $\inf_{N \geqslant 1} f_N > 0$, then Theorem 1 need not hold. For example, whenever R is sufficiently large and $w \neq 0$, we have $\frac{1}{1+w(R-1)p} < \inf_{N \geqslant 1} f_N$, so no value of N satisfies **Eq. SI.13**.

SI.2.1.2. *Variances of cooperator and defector abundances.* In state (x_C, x_D), the expected squared number of cooperators in the next generation is

$$\mathbf{E}_{(x_C, x_D)}\left[y_C^2\right] = \sum_{(y_C, y_D)} \mathbf{P}_{\text{WFB}}\Big((x_C, x_D), (y_C, y_D)\Big) m_2, \tag{SI.17}$$

where

$$m_2 = \sum_{a=0}^{y_C} \sum_{b=0}^{y_D} (a+b)^2 \binom{y_C}{a}\left(1 - \frac{u}{2}\right)^a \left(\frac{u}{2}\right)^{y_C - a} \binom{y_D}{b}\left(1 - \frac{u}{2}\right)^{y_D - b}\left(\frac{u}{2}\right)^b$$

$$= \left(1 - \frac{u}{2}\right)y_C\left(1 + \left(1 - \frac{u}{2}\right)(y_C - 1)\right) + 2\left(\frac{u}{2}\right)\left(1 - \frac{u}{2}\right)y_C y_D + \left(\frac{u}{2}\right)y_D\left(1 + \left(\frac{u}{2}\right)(y_D - 1)\right). \tag{SI.18}$$

It follows from a straightforward calculation that

$$\mathbf{E}_{(x_C, x_D)}\left[y_C^2\right] = \left[\left(1 - \frac{u}{2}\right)(x_C F_C) + \left(\frac{u}{2}\right)(x_D F_D)\right] + \left[\left(1 - \frac{u}{2}\right)(x_C F_C) + \left(\frac{u}{2}\right)(x_D F_D)\right]^2$$

$$= \mathbf{E}_{(x_C, x_D)}\left[y_C\right] + \mathbf{E}_{(x_C, x_D)}\left[y_C\right]^2. \tag{SI.19}$$

Therefore, $\mathbf{Var}_{(x_C, x_D)}\left[y_C\right] = \mathbf{E}_{(x_C, x_D)}\left[y_C\right]$, and, similarly, $\mathbf{Var}_{(x_C, x_D)}\left[y_D\right] = \mathbf{E}_{(x_C, x_D)}\left[y_D\right]$. Thus,

$$\sigma_{(x_C, x_D)}\left[y_C\right] / \mathbf{E}_{(x_C, x_D)}\left[y_C\right] = \sqrt{\mathbf{E}_{(x_C, x_D)}\left[y_C\right]}; \tag{SI.20a}$$
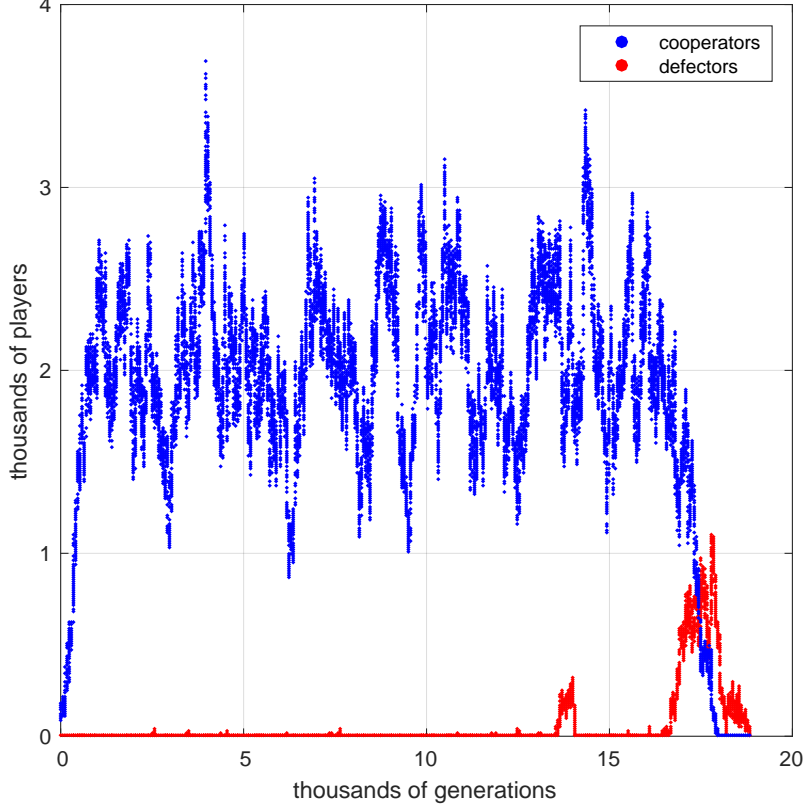
FIGURE SI.2. Mutation-selection dynamics with drastically reduced mutation rates. The parameters used here are $u = 0.00001$, $w = 0.005$, and $R = 2.0$; the baseline reproductive rate is $f_N = \max\left\{0, f_K + r\left(1 - \frac{N}{K}\right)\right\}$, where $f_K = 0.995$, $r = 0.005$, and $K = 2000$. The population is initialized with 100 cooperators and no defectors, and the cooperators then grow to reach their carrying capacity. Although small clusters of defectors are occasionally introduced through mutation, cooperators can resist invasion for a short period of time. Eventually, defectors invade and replace cooperators, and the population goes extinct due to the fact that defectors cannot survive on their own ($f_1 < 1$).

$$\sigma_{(x_C, x_D)}[y_D] / \mathbf{E}_{(x_C, x_D)}[y_D] = \sqrt{\mathbf{E}_{(x_C, x_D)}[y_D]}, \tag{SI.20b}$$

which both approach 0 as $\mathbf{E}_{(x_C, x_D)}[y_C]$ and $\mathbf{E}_{(x_C, x_D)}[y_D]$ get large.

SI.3. EXTINCTION TIME FOR BRANCHING GAMES

Let $x = (x_C/K, x_D/K)$ be normalized quantities of cooperators and defectors, where $K > 0$ parametrizes the baseline reproduction rates (i.e. the baseline reproductive rate, f_N). The "mean field" map $\phi : \mathbb{R}^2_{\geq 0} \to \mathbb{R}^2_{\geq 0}$ is defined as $\phi(x) = A(x)\, x$ where the matrix $A(x)$ is defined as

$$A(x) = \begin{pmatrix} (1-u)\, F_C(xK) & u F_D(xK) \\ u F_C(xK) & (1-u)\, F_D(xK) \end{pmatrix}. \tag{SI.21}$$

By the definitions of F_C and F_D, we know that ϕ is bounded.

We consider the normalized Markov chain $X_t^K = (C_t/K, D_t/K)$, where C_t and D_t are the number of cooperators and defectors at time t, respectively. Write $p_K(x, y) = \mathbf{P}\left[X_{t+1}^K = y \ : \ X_t^K = x\right]$ for the transition kernel. The transition probabilities are Poisson-distributed with mean given by the matrix $A(x)$. Note that $p_K(0, y) = 0$ for all $y \neq 0$ since 0 is an absorbing state.
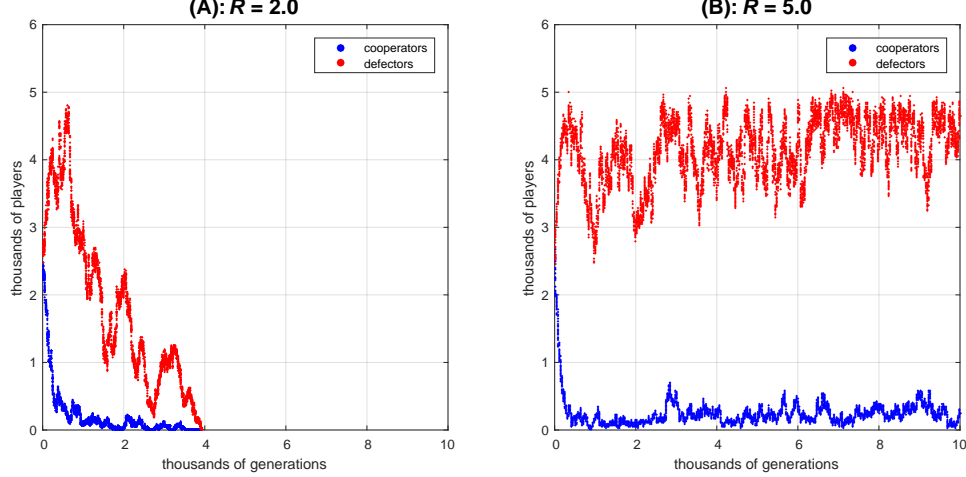
FIGURE SI.3. Mutation-selection dynamics with reduced mutation rates. The baseline reproductive rate is given by $f_N = (1 + r) \min \left\{ 1, \frac{K}{N} \right\}$, where $r = -0.001$ and $K = 5000$. In (A), everything is the same as it was in **Fig. 3**(C) apart from the mutation rate, which is $u = 0.001$ here instead of $u = 0.01$. The initial dynamics of cooperators and defectors are similar to **Fig. 3**(C), with a decline in cooperator numbers and an increase in the abundance of defectors, but ultimately the mutation rate is too small for cooperators to sustain defectors and the population goes extinct ($f_1 < 1$). The only difference between (A) and (B) is the multiplication factor for the public goods game, R, which is 2.0 in (A) and 5.0 in (B). This panel illustrates how a reduction in the mutation rate can be compensated for by an increase in the multiplication factor of the public good. This property also holds if one reduces the cost of cooperation.

A measure $\mu_K \in \Delta \left(\mathbb{R}_{\geq 0}^2 - \mathbf{0} \right)$ is a quasi-stationary distribution for p_K if there exists $\lambda_K \in (0, 1)$ such that

$$\int_{x \in \mathbb{R}_{\geq 0}^2 - \mathbf{0}} p_K (x, E) \, d\mu_K (x) = \lambda_K \mu_K (E). \tag{SI.22}$$

for all $E \subseteq \mathbb{R}_{\geq 0}^2 - \mathbf{0}$. We denote the extinction time, i.e. the time until the chain is absorbed at the state 0, by τ_K. Note that, if we start distributed according to a quasi-stationary distribution, μ_K, then the probability of being absorbed in the next step is $1 - \lambda_K$ since

$$\mathbf{P} \left[X_{t+1}^K = 0 \; : \; X_t^K \sim \mu_K \right] = \int_{x \in \mathbb{R}_{\geq 0}^2 - \mathbf{0}} p_K (x, 0) \, d\mu_K (x)$$

$$= \int_{x \in \mathbb{R}_{\geq 0}^2 - \mathbf{0}} \left(1 - p_K \left(x, \mathbb{R}_{\geq 0}^2 - \mathbf{0} \right) \right) d\mu_K (x)$$

$$= 1 - \lambda_K. \tag{SI.23}$$

Moreover, if not absorbed in the next time step, the chain remains distributed according to μ_K. Therefore, the extinction time τ_K is a geometric random variable with parameter $1 - \lambda_K$, and $\mathbf{E} \left[\tau_K \right] = 1 / (1 - \lambda_K)$, where $\mathbf{E} \left[\tau_K \right]$ denotes the expected value of τ_K when the chain is initially distributed according to μ_K.

Proposition 1. There exists $c > 0$, independent of K, such that $\mathbf{E} \left[\tau_K \right] \geq e^{cK}$.

Proof. For the model considered in the main text, ϕ has a unique fixed point, x^*, with $\phi (x^*) = x^*$. Moreover, this fixed point is an attractor. (One can show that the normalized quasi-stationary distribution μ_K converges weakly to δ_{x*} as $K \to \infty$.) Therefore, there exists $\delta > 0$ and an open set, U, containing x^* such that $N_\delta \left(\phi \left(\overline{U} \right) \right) \subseteq U$, where for $E \subseteq \mathbb{R}_{\geq 0}^2 - \mathbf{0}$, the δ-neighborhood of E is defined as

$$N_\delta (E) := \left\{ y \in \mathbb{R}_{\geq 0}^2 - \mathbf{0} \; : \; \inf_{x \in E} \| y - x \| < \delta \right\}. \tag{SI.24}$$

20

By definition of the quasi-stationary distribution, μ_K, we have

$$
\lambda_K \mu_K\left(U\right) = \int_{x\in\mathbb{R}^2_{\geqslant 0}-\mathbf{0}} p_K(x,U)\,d\mu_K(x)
$$

$$
\geqslant \int_{x\in U} p_K\left(x,U\right)\,d\mu_K\left(x\right)
$$

$$
\geqslant \mu_K\left(U\right)\inf_{x\in\overline{U}} p_K\left(x,U\right)
$$

$$
= \mu_K\left(U\right)\left(1-\sup_{x\in\overline{U}} p_K\left(x,U^c\right)\right). \tag{SI.25}
$$

For $x\in\overline{U}$, $\phi\left(x\right)\in\phi\left(\overline{U}\right)$, which implies that $N_\delta\left(\phi\left(x\right)\right)\subset N_\delta\left(\phi\left(\overline{U}\right)\right)\subset U$. Therefore,

$$
\lambda_K \geqslant 1-\sup_{x\in\overline{U}} p_K\left(x,U^c\right)
$$

$$
\geqslant 1-\sup_{x\in\overline{U}} p_K\left(x,N_\delta\left(\phi\left(x\right)\right)^c\right). \tag{SI.26}
$$

To complete the proof, we bound $p_K\left(x,N_\delta\left(\phi\left(x\right)\right)^c\right)$ via a large-deviation estimate based on the Chernoff-Cramer method. If Z is a Poisson random variable with mean $\phi\left(x\right)K$, then

$$
p_K\left(x,N_\delta\left(\phi\left(x\right)\right)^c\right) = \mathbf{P}\left[X^K_{t+1}\notin N_\delta\left(\phi\left(x\right)\right)\ :\ X^K_t = x\right]
$$

$$
= \mathbf{P}\left[\left|X^K_{t+1}-\phi\left(x\right)\right|>\delta\ :\ X^K_t = x\right]
$$

$$
= \mathbf{P}\left[\left|Z-\phi\left(x\right)K\right|>\delta K\right]. \tag{SI.27}
$$

Using Markov's inequality and the Poisson moment-generating function, we see that

$$
\mathbf{P}\left[Z>\left(\phi\left(x\right)+\delta\right)K\right]\leqslant \frac{\mathbf{E}\left[e^{tZ}\right]}{e^{t(\phi(x)+\delta)K}} = \frac{e^{\phi(x)K\left(e^t-1\right)}}{e^{t(\phi(x)+\delta)K}}. \tag{SI.28}
$$

As a function of t, the minimum of $\frac{e^{\phi(x)K\left(e^t-1\right)}}{e^{t(\phi(x)+\delta)K}}$ is at $t^* = \log\left(1+\delta/\phi\left(x\right)\right)$. Since the function $g\left(y\right):=\log\left(1+\delta/y\right)\left(y+\delta\right)-\delta$ satisfies $g\left(y\right)>0$ and $g'\left(y\right)<0$ for all $y>0$, we have

$$
\mathbf{P}\left[Z>\left(\phi\left(x\right)+\delta\right)K\right]\leqslant e^{-g(\phi(x))K}\leqslant e^{-g(m)K}, \tag{SI.29}
$$

where $m=\max_{x\in\mathbb{R}^2_{\geqslant 0}}\phi\left(x\right)$. It follows that with $c:=g\left(m\right)$,

$$
\mathbf{E}\left[\tau_K\right] = \frac{1}{1-\lambda_K}\geqslant e^{cK}, \tag{SI.30}
$$

which completes the proof. $\qquad\square$