PHYSICAL REVIEW E **106**, 014112 (2022)

When stochasticity leads to cooperation

Ian Braga [*] and Lucas Wardil
Departamento de Física, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, Minas Gerais, Brazil

(Received 4 April 2022; accepted 17 June 2022; published 12 July 2022)

The evolution of cooperation has gained more attention after Smith introduced game theory in the study of evolutionary biology. Subsequent works have extensively explained this phenomenon, consistently showing the importance of spatial structure for the evolution of cooperation. Here we analyze the effect of stochasticity on the evolution of cooperation in group-structured populations. We find a simple formula for the fixation probability of cooperators and show that cooperation can be favored by selection if a condition similar to Hamilton's rule is satisfied, which is also valid for strong selection and high migration. In fact, cooperation can be favored even in the absence of population viscosity and in the limit of an infinite number of finite-size groups. We discuss the importance of stochastic fluctuations in helping cooperation. We argue that this may be a general principle because fluctuations favoring the cooperators are often much more impactful than those favoring the defectors.

DOI: 10.1103/PhysRevE.106.014112

I. INTRODUCTION

The problem of explaining the emergence of cooperation by natural selection has been addressed recently with some microscopic models that shed much light on this apparently paradoxical phenomenon [1–10]. After all, the cooperator pays reproductive costs to benefit others in a world of fierce competition. Most authors agree that spatial restrictions delimiting the interactions of individuals can favor the evolution of cooperation, increasing population viscosity [11]. Here we argue that the stochastic nature of the evolutionary dynamics may also be one of the key features that can make cooperation thrive [1,9,12–14].

To illustrate the problem of cooperation, let us consider a colony of ancestral bacteria living in an environment with a potential resource that cannot be metabolized. Is a mutation that promotes the synthesis of a new enzyme, which the bacteria can release to break down this resource, an advantageous one? It may seem advantageous at first look. However, if the enzyme is shared as a public good, the mutant will bear the energetic cost of producing it, with the nonmutants benefiting from the resource without paying any costs. Thus, the nonmutant type will reproduce more, preventing the spread of the new cooperative mutation. The yeast *Saccharomyces cervisiae* faces this dilemma and does not produce an enzyme needed for the digestion of sucrose, relying on a defector strategy for proliferation [15]. However, if it is so challenging to have cooperation, how is the first emergence of multicellular organisms explained [16]? Or the enormous success of eusocial insects [17], with just ants corresponding to 15–20 % of terrestrial animals biomass [18]?

Based on the idea that cooperation can be advantageous because groups with more cooperators are selected, the group selection theory is advocated as one of the key mechanisms in promoting cooperation. However, it has received some discredit in the academy [19]. The main criticism is that group selection is an evolutionary force much weaker than individual selection [20]. Although subsequent works showed that group structure could favor cooperation [1,21], later critics have argued that group selection and other kinds of interacting-neighbor dynamics are just different approaches to kin selection. Moreover, it has been claimed that all these phenomena can be treated with the same mathematical framework [22]. However, these views have also been contested because misrepresentation of the concept of relatedness is often necessary [23]. In addition, in a recent work [24], the authors found that group structure can sustain cooperation under the pairwise comparison update, but cannot under the Moran process update.

Here we investigate the intrinsic stochasticity typical of evolutionary dynamics in group-structured populations. We show that the stochasticity, and not the viscosity created by the population structure, is the mechanism that drives cooperation. We use modern stochastic process methods to derive a simple analytic formula for the fixation probability of cooperators for any initial fraction and any strength of selection. We conclude that for cooperation to be advantageous, it is only necessary that an inequality similar to Hamilton's rule is satisfied: $b/c > k$, where b is the bonus of cooperation, c is its costs, and k is roughly the number of individuals in each group. More interestingly, we show that cooperation can be selected only because of the stochastic fluctuations in the system, which is observed even in the limit where offspring always migrate to different groups and in the limit of infinite populations with finite-size groups.

In the next section we define the stochastic model and derive the master equation describing the process. Then we divide the analysis into three parts. First, we look at the deterministic limit, which provides a baseline for comparing stochastic effects. Second, we analyze the regime of high migration, where we derive a simple Hamilton-like inequality

---

[*] ianmbraga@gmail.com

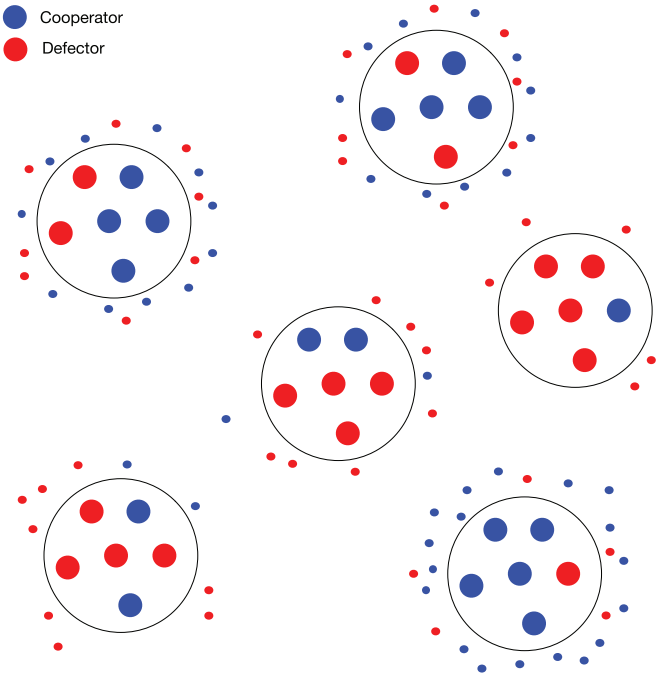

FIG. 1. Illustration of the selection process taking place in a group-structured population. The individuals live in separated groups of the same constant size and produce spores ready to flourish as soon as they find a vacant place. Eventually, spores from one group arrive in another group. So two competing evolutionary forces take place: A defector produces more spores than a cooperator of the same group, but the groups with more cooperators produce more spores than groups with more defectors. So our goal is to establish the threshold between the two competing forces and reveal the role of stochasticity.

from the analysis of the fixation probability. Third, we analyze the regime of low migration using a time-separation technique to obtain an analytical approximation for the fixation probability.

II. MODEL AND MASTER EQUATION

The analytical results in our work are based on the Moran process with N individuals [25]. The population is divided into n groups of the same size, and the reproductive fitness of the individuals is determined only by the interactions within the individual's group. In Fig. 1 we provide a schematic illustration of an example of a real process that can be approximated as a Moran process in group-structured populations.

Every time step, one individual in the entire population is randomly chosen to die. Let us suppose that the dead individual was in group i. Then another is chosen to reproduce with probability proportional to its fitness. The newborn comes from the same group i with probability proportional to μ_{ii} and from group j with probability proportional to μ_{ij}. There are two types of individuals, A and B. Let N_{Ai} and N_{Bi} be the number of individuals of type A and B in the group i, respectively, and let F_{Ai} and F_{Bi} be the fitness of each type in the group i. Note that there are only n independent variables because of the constraint $N_{Ai} + N_{Bi} = N/n$.

Let $P(\mathbf{N}, t)$ be the probability that the population is in state $\mathbf{N} = (N_{A1}, \ldots, N_{An})$ at time t. This probability obeys the master equation

$$\begin{aligned}\frac{\partial}{\partial t}P(\mathbf{N}, t) &= \sum_i T_i^+(N_{Ai} - 1)P(N_{Ai} - 1, t)\\ &\quad + \sum_i T_i^-(N_{Ai} + 1)P(N_{Ai} + 1, t)\\ &\quad - \sum_i [T_i^+(\mathbf{N}) + T_i^-(\mathbf{N})]P(\mathbf{N}, t),\end{aligned} \tag{1}$$

with transition rates given by

$$\begin{aligned}T_i^+ &= \frac{1}{Z}\frac{N_{Bi}}{N}\sum_j \mu_{ij}\frac{N_{Aj}}{N}F_{Aj},\\ T_i^- &= \frac{1}{Z}\frac{N_{Ai}}{N}\sum_j \mu_{ij}\frac{N_{Bj}}{N}F_{Bj},\end{aligned}$$

where Z is a normalization factor. The migration rates are defined as $\mu_{ij} = 1$ for $i = j$ and $\mu_{ij} = \mu$ for $i \neq j$. Note that for $\mu = 0$ a vacant place in one group is occupied only by offspring coming from the same group and for $\mu = 1$ the vacant place can be occupied by individuals from any group with equal probability. Finally, for $\mu \to \infty$ the vacant place is always occupied by an individual from another group.

III. RESULTS AND ANALYSIS

A. Deterministic limit

Defining the fraction of Ai individuals in the population by $x_i = N_{Ai}/N$ and the state vector $\mathbf{x} = (x_1, \ldots, x_n)$, we can write

$$\begin{aligned}\frac{\partial}{\partial t}P(\mathbf{x}, t) &= \sum_i T_i^+\left(x_i - \frac{1}{N}\right)P\left(x_i - \frac{1}{N}, t\right)\\ &\quad + \sum_i T_i^-\left(x_i + \frac{1}{N}\right)P\left(x_i + \frac{1}{N}, t\right)\\ &\quad - \sum_i [T_i^+(\mathbf{x}) + T_i^-(\mathbf{x})]P(\mathbf{x}, t).\end{aligned}$$

Assuming that N is large enough so that we can treat x_i as a continuum variable and $\rho(\mathbf{x}, t) = NP(\mathbf{x}, t)$ as a probability distribution, we can take the Kramers-Moyal expansion of the master equation

$$\begin{aligned}\frac{\partial}{\partial t}\rho(\mathbf{x}, t) &= \sum_i (1 - e^{-(1/N)(\partial/\partial x_i)})T_i^+(\mathbf{x})\rho(\mathbf{x}, t)\\ &\quad + \sum_i (1 - e^{(1/N)(\partial/\partial x_i)})T_i^-(\mathbf{x})\rho(\mathbf{x}, t).\end{aligned}$$

Keeping only the first two terms and rescaling the time $\tau = t/N$, we arrive at the Fokker-Plank equation

$$\begin{aligned}\frac{\partial}{\partial \tau}\rho(\mathbf{x}, \tau) &= -\sum_i \frac{\partial}{\partial x_i}[T_i^+(\mathbf{x}) - T_i^-(\mathbf{x})]\rho(\mathbf{x}, \tau)\\ &\quad + \frac{1}{2N}\sum_i \frac{\partial^2}{\partial x_i^2}[T_i^+(\mathbf{x}) + T_i^-(\mathbf{x})]\rho(\mathbf{x}, \tau),\end{aligned}$$

from which we can promptly obtain the Itô process that describes our system, that is,

$$dx_i(\tau) = (T_i^+ - T_i^-)d\tau + \sqrt{\frac{T_i^+ + T_i^-}{N}}dB_i(\tau), \tag{2}$$

where $dB_i(\tau)$ is the Brownian random variable.

We would like to see if cooperators can spread and take the environment despite having local disadvantages against defectors. Thus we take the traditional game theoretical analysis with the payoff matrix $\left(\begin{smallmatrix} b-c & -c \\ b & 0 \end{smallmatrix}\right)$, where b is the bonus of cooperation and c the associated cost.

Suppose that each group is well mixed so that the individuals interact with each other with the same probability. Then expected payoffs of types A and B in group i are

$$\begin{aligned} f_{Ai} &= (b-c)nx_i + (-c)(1-nx_i) = bnx_i - c, \\ f_{Bi} &= bnx_i + 0(1-nx_i) = bnx_i, \end{aligned} \tag{3}$$

where nx_i is the fraction of cooperators and $1 - nx_i$ the fraction of defectors in group i. The average fitness of A and B individuals in group i are given by

$$F_{Ai} = 1 + bnx_i - c, \tag{4}$$

$$F_{Bi} = 1 + bnx_i, \tag{5}$$

where we choose 1 as the base fitness without loss of generality. Clearly, we must have $c < 1$ to avoid negative probability and $b > c$ for cooperation to make sense.

Let us first look at the Itô equation in the limit $N \to \infty$ to drop the diffusion term. The deterministic set of differential equations that describes the system is then

$$\begin{aligned} dx_i(\tau) = &\frac{\gamma\mu}{1+(b-c)x}\left[\sum_j \frac{1}{n}(1+bnx_j)(x_j - x_i)\right] \\ &- \frac{\gamma c}{1+(b-c)x}\left(\frac{1}{n} - x_i\right)\sum_j \mu_{ij}x_j, \end{aligned}$$

where $x = \sum_i x_i$ and $\gamma = n/[1+\mu(n-1)]$. The analysis is trivial in this limit. The first term drives the variables near each other and becomes zero when $x_i = x_j$ for all i, j. The second term is always negative and has the effect of decreasing the value of x_i. So the only equilibrium which is a global attractor is $x_i = 0$ for all i. The defectors win.

Now we raise two issues. First, populations are always of finite size and we should analyze the role of stochastic effects carefully. Second, this deterministic limit analysis does not say what happens if $n \to \infty$. In fact, we will see that cooperation can be advantageous even in the limit $N \to \infty$, as long as we keep $N/n = \text{const}$.

B. Stochastic system with high migration

We saw that the first term of the Itô stochastic description in Eq. (2) is a deterministic drift that approximates the variables to each other. Thus, we can assume that, after a short relaxation time, we can make the approximation $x_i \approx x/n$ for all i. We use this fact to describe the system only in terms of the total fraction of cooperators x. Then we use the formula for the fixation probability of cooperators in terms of a single random variable $x \in [0, 1]$ subjected to transitions rates $T^+(x) = \sum_i T_i^+$ and $T^-(x) = \sum_i T_i^-$, with initial condition x_0 [26]. Notice that x_0 is the initial fraction of cooperators in the entire population. This probability is given by

$$\phi_A(x_0) = \frac{\int_0^{x_0} \exp\left(-\int \frac{2A(x')}{B(x')}dx'\right)dx'}{\int_0^1 \exp\left(-\int \frac{2A(x')}{B(x')}dx'\right)dx'}, \tag{6}$$

where $A(x) = T^+(x) - T^-(x)$ and $B(x) = [T^+(x) + T^-(x)]/N$.

The formula (6) is valid for any value of c. To obtain a simple threshold equation for cooperation, we take the approximation $c \ll 1$, which we show to be consistent in the parameter region of interest. Under this approximation, we obtain

$$B(x) = \frac{2x(1-x)}{N} \tag{7}$$

and

$$A(x) = \frac{\mu\gamma b/2}{1+(b-c)x}\sum_{i,j}(x_j - x_i)^2 - \frac{cx(1-x)}{1+(b-c)x}, \tag{8}$$

where we also approximate $x_i = x/n$ in the second term of $A(x)$. As we have discussed, in a deterministic description, the variables x_i should become equal as soon as the system is released. However, we cannot make the same approximation in the first term because we would get zero and completely lose the cooperation-driven effect due to the differences in x_i [note that the first term of $A(x)$ is always positive]. In a population with finite group size, it is improbable that the variables x_i always have the same value. So, depending on the magnitude of the stochastic fluctuations, the first term of $A(x)$ could counterbalance the negative rate due to the cooperation cost c.

The next step is to approximate the value of $(x_j - x_i)^2$ as $\langle (x_j - x_i)^2 \rangle$ to simplify the summation in the first term in Eq. (8). Note that we cannot neglect the cooperation cost c in Eq. (8) because we do not know how small it is compared to $\langle (x_j - x_i)^2 \rangle$. However, we know that this average quadratic distance must be attained very soon after the system is released because we have a deterministic drift trying to put all the variables together. Then we assume that the fluctuations in x_i and x_j are weak and noncorrelated so that we have

$$\begin{aligned} \langle (x_j - x_i)^2 \rangle &= \langle x_j^2 \rangle + \langle x_i^2 \rangle - 2\langle x_j x_i \rangle \\ &= \langle x_j^2 \rangle - \left(\frac{x}{n}\right)^2 + \langle x_i^2 \rangle - \left(\frac{x}{n}\right)^2 \\ &= 2\,\mathrm{Var}(x_i). \end{aligned}$$

Now we only need to calculate $\mathrm{Var}(x_i)$. Note that we start the system with the same initial fraction of cooperators in all groups, $x_i(0) = x/n$ for all i, because this metaequilibrium is reached after a small relaxation time.

To calculate $\mathrm{Var}(x_i)$, we return to the Itô stochastic description of the variables [Eq. (2)]. We take $c = 0$, $(1 + bnx_j) = (1 + bx)$ for the drift term, and $x_i = x/n$ for the diffusion therm so that our stochastic process can be approximated by

$$dx_i(\tau) = \mu\gamma\left(\frac{x}{n} - x_i\right)d\tau + \sqrt{\frac{2x(1-x)}{nN}}dB(\tau),$$

where we consider x as a constant in the small interval needed for the variable to reach its maximum variance. We also make the change of variables $x' = x/n - x_i$ to obtain

$$dx' = -\mu\gamma x' d\tau + \sqrt{\frac{2x(1-x)}{nN}}dB(\tau),$$

where we can take $-dB(\tau) = dB(\tau)$ with no change in the process. This equation can be easily solved using the Itô formula, resulting in

$$x'(\tau) = e^{-\mu\gamma\tau}\int_0^\tau e^{\mu\gamma t}\sqrt{\frac{2x(1-x)}{nN}}dB(t),$$

where $x'(0) = 0$ since $x_i(0) = x/n$. The variance of x_i is just

$$\begin{aligned}\left\langle\left(x_i(\tau) - \frac{x}{n}\right)^2\right\rangle &= \langle[x'(\tau)]^2\rangle \\ &= e^{-2\mu\gamma\tau}\frac{2x(1-x)}{nN}\int_0^\tau e^{2\mu\gamma t}\langle dB(t)^2\rangle \\ &= \frac{1}{\mu\gamma nN}x(1-x)(1-e^{-2\mu\gamma\tau}),\end{aligned}$$

where we use that $\langle dB(t)^2\rangle = dt$. As we have previously discussed, the variance becomes close to its maximum value in a very short time. So we can take $\text{Var}(x_i) \approx x(1-x)/\mu\gamma Nn$, and our estimate for the average quadratic distance of x_i and x_j is just

$$\langle(x_j - x_i)^2\rangle \approx \frac{2x(1-x)}{\mu\gamma Nn}.$$

Using these approximations, we obtain the simplified formula

$$\phi_A(x_0) = \frac{[1+(b-c)x_0]^{(\alpha+1)} - 1}{[1+(b-c)]^{(\alpha+1)} - 1}, \tag{9}$$

where $\alpha = [-b(n-1) + cN]/(b-c)$. Interestingly, the fixation probability is independent of μ. There is a balance: On one side, the increase of μ increases the cooperation-driven effect in $A(x)$; on the other side, it decreases the average quadratic distance of variables in the same proportion. The theoretical expression (9) agrees very well with the simulation results for high-migration rates ($\mu > 0.5$), as shown in Fig. 2.

Finally, to obtain the threshold that gives the cooperators a fixation probability higher than that of defectors at the same initial fraction, we have to analyze the concavity condition for the function in Eq. (9). If $\phi(x_0)$ is the fixation probability of cooperators, then $1 - \phi(x_0)$ is the fixation probability of defectors in a population starting with x_0 cooperators. Thus, the value $1 - \phi(1 - x_0)$ gives the fixation probability of defectors that we must compare to the fixation probability of cooperators $\phi(x_0)$. Hence, we must require the second derivative of the function in Eq. (9) to be negative. The second derivative is given by

$$\frac{\partial^2\phi_A(x_0)}{\partial^2 x_0} = \frac{(b-c)^2(\alpha+1)\alpha[1+(b-c)x_0]^{(\alpha-1)}}{[1+(b-c)]^{(\alpha+1)} - 1}.$$

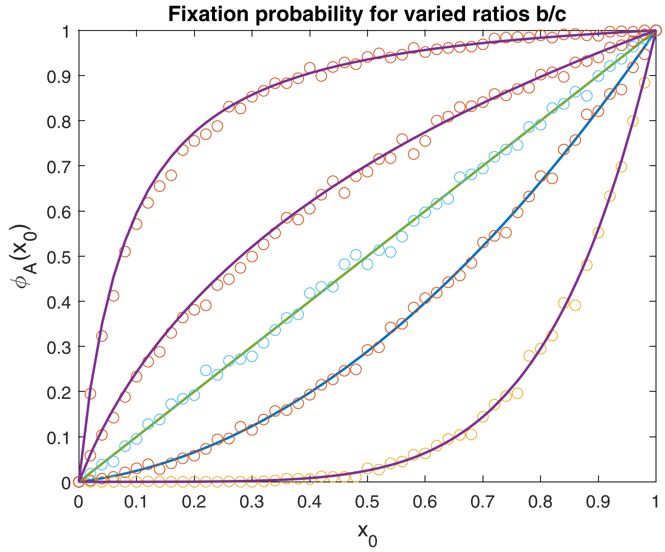

FIG. 2. Fixation probability of cooperators in the high-migration regime. The x axis is the initial fraction of cooperators in the entire population, with all groups initialized with the same fraction of cooperators. The middle curve is the neutral case $b = c = 0$, above which the cooperators have an advantage and below which the defectors have an advantage. The theoretical expression given by Eq. (9) (solid lines) agrees well with the simulation results (symbols). The parameters are $N = 400$, $n = 8$, $\mu = 1$, $b = 10$, and the values of c, from top to bottom, $c = 0.12$, 0.155, 0.2, and 0.3.

Thus, we must have

$$\frac{b}{c} > \frac{N}{n-1}. \tag{10}$$

If Eq. (10) is satisfied, we say that selection favors cooperators replacing defectors [27]. For $n = 1$, the right-hand side of Eq. (10) diverges, which is expected because cooperation has no chance in a single well-mixed group. For large N and n, this condition can be approximated by $b/c > k$, where $k = N/n$ is the number of individuals in the same group. Thus we have just recovered a condition similar to Hamilton's rule [28]. Astonishingly, in the limit $N, n \to \infty$ with $k = \text{const}$, cooperators have probability one of taking the environment if the condition in Eq. (10) is satisfied.

C. Stochastic system in low migration

In the low-migration regime ($\mu < 0.5$), the formula (9) is not accurate. In this case, the deterministic drift that approximates the variables to each other is weak and the approximation $x_i \approx x/n$ is no longer reasonable. However, if the migration between groups is rare, the dynamics can be analyzed using a timescale separation technique [29]. If mutation is rare, all groups reach fixation before the next migration occurs. Thus, what is really observed in a slow timescale is that each group is composed of a single type and that the number of groups of type A is a discrete variable jumping in the state space $\{0, 1, \ldots, n\}$.

The fixation probability of type A in the Moran process in a well-mixed population of size N [30], starting with N_A

individuals of type A, is given by the formula

$$\phi(N_A) = \frac{1 + \sum_{j=1}^{N_A-1} \prod_{k=1}^{j} \gamma(k)}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} \gamma(k)},$$

where

$$\gamma(k) = \frac{T^-(k)}{T^+(k)}$$

is defined in terms of the transition rates, which are given by

$$T^+(N_A) = \frac{1}{Z}\frac{N_A}{N}\left(1 - \frac{N_A}{N}\right)F_A(N_A),$$
$$T^-(N_A) = \frac{1}{Z}\frac{N_A}{N}\left(1 - \frac{N_A}{N}\right)F_B(N_B). \quad (11)$$

Recall that in our study the fitness is given by $F_A(N_A) = 1 + bN_A/N - c$ and $F_B(N_A) = 1 + bN_A/N$. In a well-mixed population, the fixation probability of a cooperator is never higher than the fixation probability of a defector (compared with the same initial fraction of individuals), no matter how small the cooperation cost c is. In fact, it becomes exponentially small with the increase of the population size. However, as we showed for the high-migration regime, the cooperators may have an overall advantage if we consider the competition between groups.

In the slow timescale, each group is composed of a single type. Let i be the number of groups of type-A individuals and $n - i$ the number of groups of type-B individuals. Let $\Phi_A(i)$ be the fixation probability of type A in the whole population if the system starts with i groups of type A. There are two absorbing states: $i = 0$ and $i = n$. Because the migration probability is small, all groups are always homogeneous when a new migration happens. The probability that this migration leads to an increase in one in the number of groups of type A ($i \to i + 1$) is given by

$$\mathcal{T}^+ = \frac{1}{Z}\frac{i}{n}\frac{n-i}{n}\frac{F_A(N)}{F_A(N) + F_B(0)}\phi(1),$$

where $F_A(N) = 1 + b - c$ and $F_B(0) = 1$ are the fitness of the individuals in the type-A and type-B groups, respectively. Similarly, the probability that the migration increases the number of groups of type B ($i \to i - 1$) is given by

$$\mathcal{T}^- = \frac{1}{Z}\frac{i}{n}\frac{n-i}{n}\frac{F_B(0)}{F_A(N) + F_B(0)}[1 - \phi(N - 1)].$$

Note that the Markov chain in the state space of homogeneous groups is analogous to the Moran process in a well-mixed population. Therefore, the fixation probability $\Phi_A(i)$ is given by

$$\Phi_A(i) = \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^{j} \Gamma(k)}{1 + \sum_{j=1}^{n-1} \prod_{k=1}^{j} \Gamma(k)},$$

where $\Gamma(k) = \mathcal{T}^-(k)/\mathcal{T}^+(k)$.

Finally, the fixation probability of cooperators if one of the n groups starts with N_A cooperators is just

$$\phi(N_A)\Phi_A(1), \quad (12)$$

which is the product of the fixation probability of N_A cooperators in one well-mixed group and the fixation probability of one type-A group.

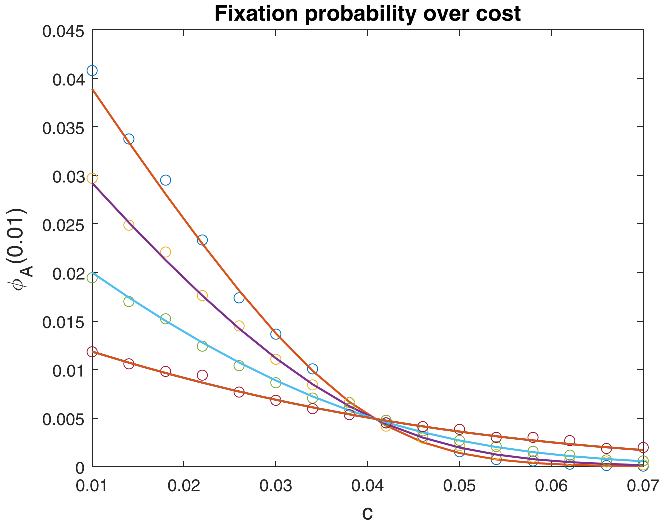

FIG. 3. Fixation probability of cooperators in the low-migration regime. The x axis is the cooperation cost c. The initial fraction of cooperators is 1%, all placed in the same group. Notice that there is an intercept point of the curves occurring approximately at $c^* = 0.04 = b/k$. The parameters are $k = 50$, $b = 2$, $\mu = 0.001$, $b = 2$, and, from top to bottom (on the left side), the number of groups $n =$ 8, 6, 4, and 2.

In Fig. 3 we compare the simulations and the analytical formula (12) for different values of the cost c. We considered a small initial fraction of 1% of cooperators in the total population, all cooperators starting in the same group. If the costs are small, cooperation can be advantageous for any quantity of groups ($n > 1$), fixating more than 1% of the time.

In this low-migration scenario, stochasticity is only responsible for the initial fluctuation that fixates cooperation in one of the groups. After this initial lucky event, the mechanism that promotes cooperation is the protection from the defectors due to the differences in the group payoffs and population timescales. Recall that cooperators always have more trouble reaching fixation in a single well-mixed group than the defectors have. However, once the inside-group dynamics has stabilized, if at least one group of type A is formed, it generates more migrants because its individuals have the maximum fitness. Nevertheless, it is still stochasticity that prevents all groups from being always initially dominated by defectors, an indispensable condition for the temporal protection of defectors to work.

Notice in Fig. 3 that all curves intercept at a point c^*. For $c < c^*$, having more groups in the population is good for the cooperators (keeping the group size constant), but for $c > c^*$ the cooperators are better off if the population has few groups. Interestingly, this value is given approximately by $c^* \approx b/k$, where $k = N/n$ is the size of the group. The explanation for the result is simple. A single cooperator in a group of defectors pays a cost c to provide the group a benefit b, which all k individuals share. The average payoff of a cooperator is then given by $b/k - c = c^* - c$, which is positive if $c < c^*$ and negative if $c > c^*$. Thus, if $c < c^*$, the cooperator does better than any individual in an all-defector group and adding a new all-defector group will only enhance the cooperator advantage relative to the total population. If $c > c^*$, adding

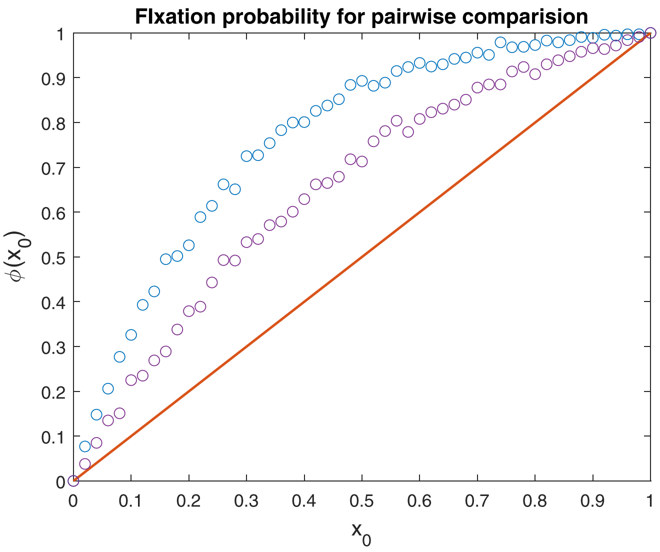

FIG. 4. Simulations for pairwise comparison update. The x axis is the initial fraction of cooperators and the y axis is the fixation probability. The parameters are $N = 200$, $n = 4$, $\mu = 1$, $b = 2$, and the values of c, from top to bottom, $c = 0.01$ and 0.02. The solid curve is the neutral case. The cooperators are advantageous for small costly cooperation for any initial condition.

a new all-defector group decreases the relative advantage of the cooperator in the population and it is better to have fewer groups.

IV. OTHER RULE UPDATES

The Moran process in group-structured populations is a model that yields simple analytical expressions for the fixation probability. We also verified through simulations that other update rules yield the same results, showing that small, costly cooperation can be favored by selection.

A. Pairwise comparison

In the pairwise comparison update, a randomly selected individual is replaced by a clonal offspring of another randomly picked individual with a probability that depends on the fitness difference between the model and the focal individual. The Fermi function is often adopted as the probability of strategy replacement, with the first chosen individual X adopting the strategy of the second chosen individual Y with probability

$$p_{X\to Y} = \frac{1}{1+e^{F_X - F_Y}},$$

where F_X and F_Y are the fitness of individuals X and Y, respectively. In our model, the transition rates for this update rule becomes

$$T_i^+ = \frac{n}{1+\mu(n-1)}\left(\frac{1}{n} - x_i\right)\sum_j \frac{\mu_{ij}x_j}{1+e^{bn(x_i-x_j)+c}},$$

$$T_i^- = \frac{n}{1+\mu(n-1)}x_i\sum_j \frac{\mu_{ij}(\frac{1}{n} - x_j)}{1+e^{bn(x_i-x_j)-c}}.$$

The simulation results are shown in Fig. 4.

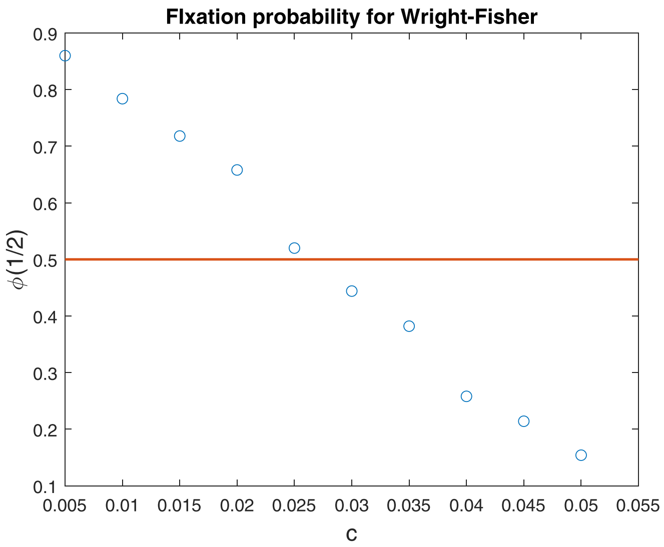

FIG. 5. Simulations for Wright-Fisher update. The x axis is the cost of cooperation and the y axis is the fixation probability for $x_0 = 1/2$. The parameters are $N = 150$, $n = 3$, and $b = 2$. The cooperators are advantageous for small costly cooperation. In particular, for these parameters, they dominate more than half of the times when $c < 0.025$.

B. Wright-Fisher process

In the Wright-Fisher update, every time step, all individuals of the current generation are replaced by their offspring, which are randomly placed in the groups. If the population is in state $(N_{A1}, \dots, N_{An})$ at time t, any vacant place in the new generation t' will have a probability

$$p = \frac{1}{1+(b-c)x}\sum_i x_i(1+bx_i-c)$$

of being filled by a cooperator and a probability $q = 1 - p$ of begin filled by a defector. Therefore, the probability of any group i having N'_{Ai} cooperators in the next generation t' is given by the binomial distribution

$$P(N'_{Ai}) = \binom{N/n}{N/n - N'_{Ai}} p^{N'_{Ai}} q^{N/n - N'_{Ai}}.$$

The simulations of this dynamics are shown in Fig. 5. As before, small costly cooperation favors cooperation.

V. DISCUSSION

Finally, let us compare our results with previous works. First, our analysis sheds light on the discussion between group selection and kin selection. Although both approaches may sometimes give the same threshold for cooperation emergence [4,22], the mechanisms are different. Cooperation driven by kin selection is due to interacting with close relatives preferentially. Cooperators end up interacting more often with other cooperators, preventing the exploitation by defectors. It is understandable to view the cooperation induced by interacting-neighbor models on graphs as an instance of kin selection [5,31], since cooperators interact more often with cooperators as a consequence of population viscosity [3]. However, in our model, it is not the viscosity of the population that promotes cooperation, but the competition between the groups, where by competition we mean the groups competing to send more migrants to occupy new vacant places. In fact, we showed that the threshold for cooperation does not depend on the migration rate for $\mu > 0.5$ and cooperators can still be advantageous in the limit $\mu \to \infty$, where a newborn individual in one group always comes from another group.

The role of the stochasticity in our model is even more evident if we look at the deterministic model, making $dB_i = 0$ in Eq. (2). In this scenario, cooperation cannot be sustained for any population size N and ratio b/c. After a relaxation time, the groups have the same proportion of cooperators, stopping the competition between them. Because cooperation induced by viscosity can be advantageous even in deterministic models [11,32,33], the stochastic effect may play a significant role in promoting cooperation. Also, it is commonly said that the rule $b/c > k$ for the evolution of cooperation in graphs is a consequence of the linearity introduced by weak selection [23,34]. In our model, for high migration, this simple rule is extended for any strength of selection.

VI. CONCLUSION

To sum up, our analysis suggests that, in addition to positive correlation in the interaction between cooperators caused by spatial structure, the stochastic nature of the evolutionary dynamics can be a core mechanism for cooperation success. In [1] Fletcher and Zwick analyzed a similar version of our model using the Wright-Fisher update and, using inclusive fitness analyses, found a similar threshold $b/c > k$. They provided an intuitive explanation of stochasticity's role in favoring cooperation. In their words, "[t]he groups that are by chance initially dominated by altruists grow larger compared to other groups and even though the fraction of altruists declines in these groups, the absolute number of altruists poised to benefit other altruists in a subsequent generation increases." The same holds in our model, where cooperation can only be advantageous because some groups have more cooperators than others due to pure fluctuations, which may increase the overall fraction of cooperators in the population due to the indirect competition between groups. Hence, we claim that stochasticity is a key mechanism for promoting cooperation in populations structured in groups, since the fluctuations that happen to favor cooperators tend to have much more impact than those that favor defectors.

ACKNOWLEDGMENT

The authors thank the Brazilian agencies CNPq and CAPES.

[1] J. A. Fletcher and M. Zwick, J. Theor. Biol. **228**, 303 (2004).

[2] Q. Su, A. McAvoy, L. Wang, and M. A. Nowak, Proc. Natl. Acad. Sci. USA **116**, 25398 (2019).

[3] H. Ohtsuki, C. Hauert, E. Lieberman, and M. A. Nowak, Nature (London) **441**, 502 (2006).

[4] A. Grafen, J. Evol. Biol. **20**, 2278 (2007).

[5] G. Szabo and G. Fath, Phys. Rep. **446**, 97 (2007).

[6] L. Lehmann, L. Keller, and D. J. T. Sumpter, J. Evol. Biol. **20**, 2284 (2007).

[7] M. A. Nowak, C. E. Tarnita, and T. Antal, Philos. Trans. R Soc. Lond. B Biol. Sci. **365**, 19 (2010).

[8] P. M. Altrock, A. Traulsen, and M. A. Nowak, Phys. Rev. E **95**, 022407 (2017).

[9] M. Assaf, M. Mobilia, and E. Roberts, Phys. Rev. Lett. **111**, 238101 (2013).

[10] Q. Su, L. Aming, L. Wang, and H. E. Stanley, Proc. R. Soc. B **286**, 20190041 (2019).

[11] J. Mitteldorf and D. S. Wilson, J. Theor. Biol. **204**, 481 (2000).

[12] J. H. Gillespie, Genetics **76**, 601 (1974).

[13] B. Houchmandzadeh and M. Vallade, BMC Evol. Biol. **12**, 61 (2012).

[14] B. Houchmandzadeh, Bio. Systems **127**, 60 (2015).

[15] D. Greig and M. Travisano, Proc. R. Soc. B **271**, S25 (2004).

[16] S. A. West, R. M. Fisher, A. Gardner, and E. T. Kiers, Proc. Natl. Acad. Sci. USA **112**, 10112 (2015).

[17] E. O. Wilson and B. Holldobler, Proc. Natl. Acad. Sci. USA **102**, 13367 (2005).

[18] T. R. Schultz, Proc. Natl. Acad. Sci. USA **97**, 14028 (2000).

[19] D. S. Wilson and E. O. Wilson, Q. Rev. Biol. **82**, 327 (2007).

[20] G. C. Williams, *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Princeton University Press, Princeton, 1996).

[21] D. S. Wilson, Proc. Natl. Acad. Sci. USA **72**, 143 (1975).

[22] L. Lehmann, L. Keller, S. West, and D. Roze, Proc. Natl. Acad. Sci. USA **104**, 6736 (2007).

[23] M. A. Nowak, C. E. Tarnita, and E. O. Wilson, Nature **466**, 1057 (2010).

[24] C. Hauert and L. Imhof, J. Theor. Biol. **299**, 106 (2012).

[25] A. Traulsen, J. C. Claussen, and C. Hauert, Phys. Rev. Lett. **95**, 238701 (2005).

[26] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 1985).

[27] C. Taylor, D. Fudenberg, A. Sasaki, and M. A. Nowak, Bull. Math. Biol. **66**, 1621 (2004).

[28] W. D. Hamilton, J. Theo. Bio. **7**, 1 (1964).

[29] S. Yagoobi and A. Traulsen, Sci. Rep. **11**, 17979 (2021).

[30] M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, Cambridge, 2006).

[31] P. D. Taylor, T. Day, and G. Wild, Nature (London) **447**, 469 (2007).

[32] M. A. Nowak and R. M. May, Nature (London) **359**, 826 (1992).

[33] T. Pfeiffer, S. Schuster, and S. Bonhoeffer, Science **292**, 504 (2001).

[34] C. E. Tarnita, H. Ohtsuki, T. Antal, F. Fu, and M. A. Nowak, J. Theor. Biol. **259**, 570 (2009).