Replicator population dynamics of group interactions: Broken symmetry, thresholds for metastability, and macroscopic behavior

Emmanuel Artiges,[1] Carlos Gracia-Lázaro ●,[2,3] Luis Mario Floría,[2,4] and Yamir Moreno[2,3,5]

[1]*École Normale Supérieure de Lyon, Université Claude Bernard Lyon I, 69342 Lyon Cedex 07, France*
[2]*Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza 50009, Spain*
[3]*Departamento de Física Teórica. University of Zaragoza, Zaragoza 50009, Spain*
[4]*Departamento de Física de la Materia Condensada, University of Zaragoza, Zaragoza 50009, Spain*
[5]*ISI Foundation, Turin, Italy*

The effect of group structure on cooperative behavior is not well understood. In this paper, we study the dynamics of a public goods game involving n-agent interactions. In the proposed setup, the population is organized into groups. We associate the individual fitness to group performance, while the evolutionary dynamics takes place globally. We derive analytical expressions and show that the model exhibits several fixed points, including the symmetric homogeneous states of total cooperation and total defection, which are unstable and stable, respectively. Interestingly, even if both individual and group levels are organized as well-mixed populations, the dynamics displays intermediate values of cooperation under the replicator dynamics. Namely, as soon as one of the groups, at least, is fully cooperative, intermediary fixed points appear for the rest of the groups. In addition to the analytical approach, we have performed numerical simulations that reproduce the internal fixed points obtained theoretically, showing coexisting intermediate levels of cooperation. Potential implications of these results in terms of group selection and the role of social norms are also discussed.

I. INTRODUCTION

Understanding how cooperative behavior emerges in different contexts remains an outstanding question in modern evolutionary sciences [1,2]. The challenge posed by the observation of cooperation, when selfish behavior provides higher fitness, has been studied in many different contexts, from biology [3,4] to economics [5] and sociology [6]. Several mechanisms have been proposed to explain this "cooperation dilemma." Among these, direct [7] and indirect [8,9] reciprocity rely on the idea that cooperative behavior may be favored by the likelihood of future interactions. However, network reciprocity [10] is rooted on the assortative effects of the topology of individual connections. Other proposals include kin selection, which refers to cooperation as favoring the reproductive success of an agent's relatives, even at a cost of survival or reproduction of the individual [11], or group selection, which involves competing groups of individuals. They, however, have got entangled in the old debate in evolution about multilevel selection, which have left open the question of the effective replicator: the individual, the group, the clade, the selfish gene, etc. [12–14].

Although most of the theoretical [15,16] and experimental [17–19] studies on cooperation have focused on pairwise interactions, many biological [3], social [5], or economic [6] systems which are interesting from the perspective of cooperation involve n-agent interactions. The public goods game (PGG) is perhaps one of the simplest, and most studied, "group interaction" [20] scenarios. In this game, while only cooperators contribute to the common good, both cooperators

and defectors benefit from it, that is, defectors are free-riders, social parasites. In the classical formulation of the PGG, for a constant individual cooperation cost, and a linear benefit function, defection is the rational choice and constitutes the only Nash equilibrium of the game. However, interior Nash equilibria can be found for nonlinear convex benefit functions [21]. Other solutions proposed for the resilience of cooperation in the evolutionary dynamics of PGG include structured populations [22,23] or information exchange [24].

In this paper, we investigate in detail how the strategies of the PGG spread over the population of players under maybe the simplest possible set of assumptions on the structure of agent contacts: Agents are partitioned into groups, each group constituting a fully connected subpopulation within which its members exploit a common good, the individual fitness of each agent being a simple function of the abundance of cooperators in his group. To implement Darwinian competition for strategy spreading we use the "well-mixed" myopic replicator dynamics [25]; i.e., "An agent imitates another, randomly chosen from the whole population, with a probability proportional to the fitness difference, if positive." The model is presented in Sec. II.

In Sec. III we study the mesoscale description leading to an m-dimensional Markov process describing the evolution of the fraction of cooperators in each of the m competing groups. The complete analysis of this dynamical system is greatly simplified by the use of symmetry arguments, provided the competing groups are equally sized (a generalization to unequal-sized groups is studied in Appendix C). This method provides metastable macroscopic states, showing
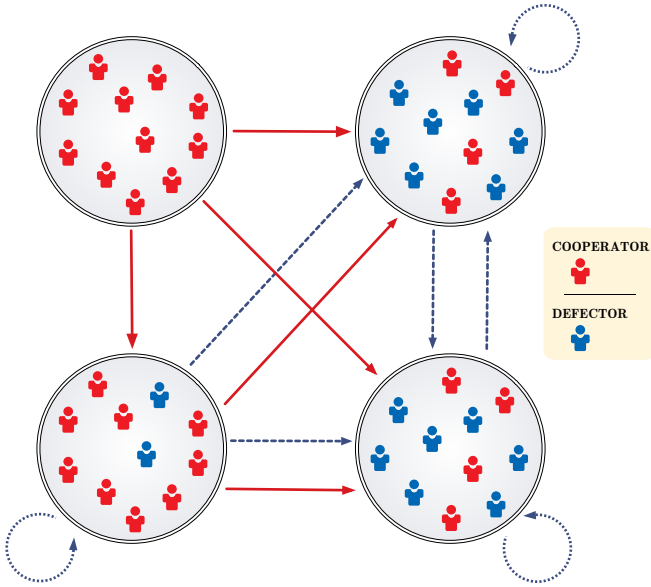
FIG. 1. Structure of the model. Strategist agents are disposed in groups and can either cooperate or defect. Agents' payoffs are proportional to the number of cooperators in their group, cooperators having to pay an extra cost. Although agents obtain their payoffs from their group, imitation takes place from any group. In the diagram, solid red arrows represent the cooperation invasion flows, and blue dashed arrows the defection flows. See the text for further details.

intermediate levels of cooperation, for a subset of initial conditions, namely, when a group is composed exclusively by cooperators. Section IV is devoted to present results agents' stochastic simulations, which show a long term behavior in which the group cooperation values fluctuate around some well defined values, accurately predicted by the Markov dynamics analyzed in the previous section. Finally, in Sec. V we discuss the implications of the model in terms of social norms, group selection, and sustainability of the common good, and present our conclusions.

II. THE MODEL

This is a simple model of evolutionary game dynamics, where a population of strategists play a PGG, from where they earn their reproductive power, i.e., their fitness. A population of $N = m \times n$ agents is divided into m groups, each one with n agents (Fig. 1). Inside each group, the agents play a public goods game with two possible strategies: cooperate (C) and defect (D). Let us denote by f_g^C and f_g^D the payoff earned, respectively, by a cooperator and a defector in the group g ($=1, \ldots, m$). If c_g is the number of cooperators in group g and the individual contribution to the public good is fixed to 1, then these payoffs are, in the standard linear PGG,

$$f_g^C = \frac{rc_g}{n} - 1, \quad f_g^D = \frac{rc_g}{n}, \tag{1}$$

where the slope $r > 1$ of the payoff is often called "synergy factor." The fitness of an individual is assumed to be proportional to his payoff. Equations (1) can be rewritten as

$$f_g^C = rp_g - 1, \quad f_g^D = rp_g, \tag{2}$$

where p_g is the fraction of cooperator agents in group g.

To implement the Darwinian competition for strategic reproduction/spread, we choose the discrete version of the myopic replicator dynamics [25,26]. On one hand, its continuum limit for pairwise games is, straightaway [25], the well-studied replicator equation [27]. On the other hand, its threshold character, see Eq. (3) below, i.e., "never change state by imitation of a less fit agent," turns out to be a source of simplicity in the analysis, as we will see below.

The idea of this model is pretty clear, namely, fitness comes from the group, while imitation takes place anywhere. Specifically, at each time step, which represents one generation of the discrete evolutionary time, all the agents play a one-shot PGG and obtain a payoff. After that, the individuals synchronously update their strategies in the following way: each agent i compares its payoff with that of a random agent j chosen equiprobably from any group, including its own. Subsequently, if agent j has a lower payoff than agent i, it keeps his/her strategy, while if it is higher i imitates j's strategy with a probability $\Pi^{j \to i}$ proportional to the payoff difference:

$$\Pi^{j \to i} = \frac{f^j - f^i}{\Delta f_{\max}} \theta(f^j - f^i), \tag{3}$$

where θ is the Heaviside step function ($\theta(y) = 1$ for $y > 0$ and $\theta(y) = 0$ for $y \leqslant 0$), and $\Delta f_{\max}$ is defined as the maximal possible difference in payoff. Note that, in this model, the fitness is given directly by the payoff, both quantities being equal. The highest payoff is the one of a single defector in a group of cooperators or the one of a cooperator in a fully cooperative group (depending on the value of r), if all other groups are 100% defectors. Besides, if all groups consist of cooperators only, except one formed by defectors but one cooperator, then the latter has the lowest payoff in the whole population. Additionally, we have that

$$\Delta f_{\max} = f_i^D(c_g = n - 1) - f_j^C(c_h = 1), \text{ if } r \leqslant n,$$
$$\Delta f_{\max} = f_i^C(c_g = n) - f_j^C(c_h = 1), \text{ if } r > n, \tag{4}$$

where c_g (respectively, c_h) stands for the number of cooperators in the group to which i (j) belongs. The value of $\Delta f_{\max}$ determines the characteristic time scale. Note that for the simplest partition, where the game is played by all agents altogether, i.e., $m = 1$, one realizes easily from Eq. (1) that the extinction of the cooperative strategy is the only possible evolutionary outcome. Free-riding is an unbeatable strategy.

III. MARKOV PHASE SPACE ANALYSIS

In this section we will consider equal-sized groups of n individuals. As there is no networked structure of agents interactions, neither regarding imitation rule nor inside groups regarding payoff earning, the fraction of cooperators in each group p_g ($g = 1, \ldots, m$) specify the relevant instantaneous description of the system's state of cooperation. However, note that if the term microstate refers to the specification of the particular state of each agent, and these are distinguishable individuals, then each particular set of values of p_g represents a different number of microstates.
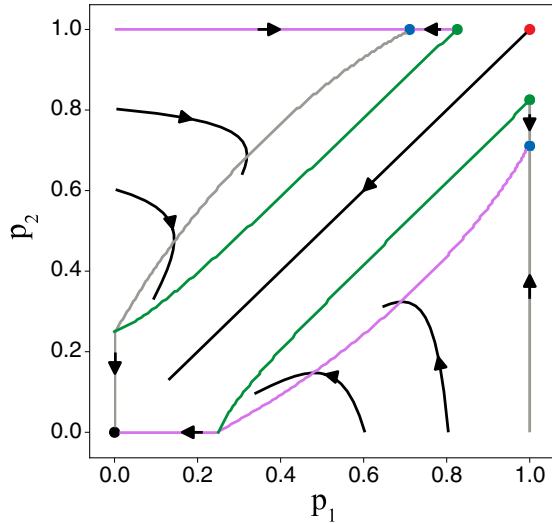
FIG. 2. Phase portrait for the two-groups case. A red dot indicates an unstable fixed point, a black dot a stable one, and a blue dot indicates a saddle point. The green lines correspond to the $f_1^C = f_2^D$ and $f_1^D = f_2^C$ singularities. The green dots correspond to the points $(1, p_2^{\text{th}})$ and $(p_1^{\text{th}}, 2)$. Gray (respectively, purple) lines correspond to nullclines $dp_1 = 0$ ($dp_2 = 0$), while arrows represent the trajectories. Here, $r = 4$. See the text for further details.

The phase space is a grid, with lattice constant $1/n$, over the m-dimensional unit hypercube. The stochastic population dynamics introduced in Sec. II defines a Markov process in this phase space, the model dynamics.

We show below the analysis of the model dynamics for the simplest cases, say $m = 2$ (Sec. III A) and 3 (Sec. III B), through a detailed geometrical investigation. The arguments used in the analysis of these explicitly solvable cases, are however easily seen to be valid for general values of the number m of groups involved. We take advantage of the simplicity that permutation symmetry considerations introduce in the analysis of the general m case in Sec. III C.

In Sec. IV stochastic simulation results are interpreted in the light of the previous nonlinear phase space analysis.

A. Two groups

Here, our phase space is the unit square. For equal-sized groups, the invariance by interchange of group label (meaning that nothing at all changes if labels 1 and 2 are interchanged everywhere) allows us to restrict attention to the triangle $0 \leqslant p_2 \leqslant p_1 \leqslant 1$. It is simple to realize that the flow points outwards nowhere on the triangle boundary. Notwithstanding this invariance, we will show in the illustrating Fig. 2 a "full" phase space portrait where this symmetry, at a first glance, can be easily acknowledged.

The stochastic dynamics defined in the previous section above for the agents' state evolution gives the following probabilities for the four possible relevant outcomes of the time step, say, increase or decrease of the number of cooperators in either group. Let P_i^+ be the probability of the transition $p_i \rightarrow p_i + 1/n$, $(i = 1, 2)$, and P_i^- the probability the probability of the transition $p_i \rightarrow p_i - 1/n$. These probabilities are

functions of the fractions p_i of cooperators:

$$P_1^+ = 0,$$

$$P_1^- = \frac{p_1}{\tau}\big[(1 - p_1) + (1 - p_2)\big(f_2^D - f_1^C\big)\ \theta\big(f_2^D - f_1^C\big)\big],$$

$$P_2^+ = \frac{1 - p_2}{\tau}\ p_1\big(f_1^C - f_2^D\big)\ \theta\big(f_1^C - f_2^D\big),$$

$$P_2^- = \frac{p_2}{\tau}\big[(1 - p_1)\big(f_1^D - f_2^C\big) + (1 - p_2)\big], \tag{5}$$

where τ is the characteristic time scale (here, $\tau = 2\Delta f_{\max}$). In the previous formulas, we have considered that $f_i^D - f_i^C = 1$. Note that, by imposing the condition $p_2 \leqslant p_1$, we are introducing an asymmetry between P_1^+ and P_2^+ also between P_1^- and P_2^-. The flow on this (p_1, p_2) unit square is $(i = 1, 2)$

$$dp_i = \frac{1}{n}(P_i^+ - P_i^-). \tag{6}$$

First we locate the nullclines, $dp_i = 0$, on the triangle.

1. $dp_1 = 0$. This locus includes the corners $(0,0)$ and $(1,1)$, and the segment of the edge $p_1 = 1$ of p_2 values for which $f_1^C > f_2^D$. The conditions $f_1^C = f_2^D$ and $p_1 = 1$ determine the upper bound, p_2^{th}, of this branch of nullcline:

$$p_2^{\text{th}} = \frac{\sqrt{1 + 4r(r - 1)} - 1}{2r}. \tag{7}$$

Note that for $p_2 > p_2^{\text{th}}$, $dp_1 < 0$, and the flow on the edge points inwards, while for $0 \leqslant p_2 < p_2^{\text{th}}$, the Heaviside function vanishes and the flow is colinear to the edge (a cooperator in group 1 does not imitate defectors in group 2).

2. $dp_2 = 0$. This includes the corners $(0,0)$ and $(1,1)$, and two branches. The first branch is the segment $0 \leqslant p_1 \leqslant 1/r$ on the edge $p_2 = 0$. The second branch is interior to the simplex, and its graph connects the points $(1/r, 0)$ and $(1, p_2^*)$, where p_2^* is explicitly computed as

$$p_2^* = \frac{r - 1}{r + 1}. \tag{8}$$

We see that this nullcline shows a singularity at $p_1 = 1/r$.

Due to the Heaviside θ functions in Eq. (5), there is a line of singularities of the flow field, namely, the intersection of the locus $f_1^C = f_2^D$ with the simplex. This is a curve connecting the points $(1/r, 0)$ and $(1, p_2^{\text{th}})$, where the constant-dp_2 isoclines show a singular behavior, similar to that of the $dp_2 = 0$ nullcline, that we have seen above.

The stationary states of the phase space flow have to be in the intersection of the nullclines. Thus, there are three fixed points $(0,0)$, $(1,1)$, and $(1, p_2^*)$. To determine the stability properties of these fixed points, one should first compute the flow's Jacobian matrices at them, whose spectral decompositions inform us on their stability against perturbations in the linear regime.

The (i, j) Jacobian matrix's element, expresses how a variation in p_j, in the linear approximation, modifies dp_i, i.e., it is

$$J_{ij} = \frac{\partial dp_i}{\partial p_j},$$

and the results of the spectral analysis of the Jacobian matrix at the three fixed points (see Appendix A for details) can be summarized as follows:

(0,0) Both eigenvalues are negative, as any fluctuation is damped out. Thus, complete defection is a local attractor. In fact, it is the global attractor for any initial condition in the interior of the simplex.

(1,1) Both eigenvalues are positive, meaning that cooperation in both groups is a repeller fixed point. Any perturbation (the appearance of a defector in either group) is amplified.

$(1, p_2^*)$ This is always a saddle point. Its stable manifold is the branch of the $dp_1 = 0$ nullcline, the segment $0 \leqslant p_2 < p_2^{\text{th}}$ on the edge. The unstable linear manifold is tangent to the interior branch of the $dp_2 = 0$ nullcline. In Fig. 2, the saddle $(p_1 = 1, \ p_2 = p_2^*)$ is represented by a blue dot, and the upper point of the $dp_1 = 0$ nullcline, $(p_1 = 1, \ p_2 = p_2^{\text{th}})$ by a green dot.

Figure 2 displays the phase portrait for $r = 2$, where previous results can be checked by simple inspection.

B. Three groups

While we analyze here the case $m = 3$, we also keep an eye on general m values, because several conclusions from this analysis are easily seen to remain valid for an arbitrary large number of groups in the system.

The set of microstates is the unit cube. One easily realizes that the fully defective corner $(0,0,0)$ is a fixed point of the dynamics. Also, it is easily seen that any small increase from zero in the fraction of cooperators in one or more groups, induces a restoring flow. Full defection is an absorbing state, an attractor. However, the other fully symmetric corner, the full cooperation corner, $(1,1,1)$ is also a fixed point, but it is unstable against defective fluctuations in one or more groups, and then is a repeller. The main diagonal connecting both, i.e., the set of fully symmetric states, is a flow trajectory of strictly decreasing value of cooperation in every group. Indeed, after a little reflection, all this is true *mutatis mutandi* for any value of $m \geqslant 2$. In any dimension m of the phase space, the set of fully symmetric microstates is invariant, and moreover, any small fluctuation orthogonal to it induces a restoring flow.

The remaining six corners of the unit cube are not fixed points, for at least one group is full defective, and at least another one is full cooperative, and then a flow of increasing fraction of cooperators in the full defective group is ensured. Also this argument applies independently of the value of m.

Now we look for eventual fixed points on the twelve edges of the unit cube. Due to the symmetry by labels interchange, they are grouped into three classes of equivalence, which correspond to invariant subsets under symmetry transformations:

1. The three axes are an invariant set under symmetry transformations. To fix ideas, think of the p_3 axis, at an abscissa $0 < p_3$. The components of the flow orthogonal to this axis are both positive, thus we conclude that the are no fixed points on the axes other than the attractor at origin. Note that this was already clear for the two-groups case; it is easy to realize that it is true for any number m of groups involved.

2. The six cube edges that are neither adjacent to the origin nor to the $(1,1,1)$ corner [e.g., the segment $(0, p_2, 1)$] form

the second equivalence class. In these microstates (points in these axes) there is one fully defective group and the flow is nonnull due to the zero payoff of its defectors. Let us note that this simple consideration leads also to the conclusion that the three faces of the unit cube adjacent to origin cannot have on them a fixed point other than the origin.

3. The three cube edges adjacent to the full cooperation corner, say the segments $(p_1, 1, 1)$, $(1, p_2, 1)$ and $(1, 1, p_3)$, form the third class of edges. To fix ideas, think of, e.g., $(1, p_2, 1)$. For values of p_2 large enough, defectors in group 2 have larger payoff than cooperators in the fully cooperative groups 1 and 3, and there, the flow points toward the interior of the unit cube. If we denote by p^{th} the value of p_2 for which $f_2^D = f_1^C = f_3^C$, and $p_1 = p_3 = 1$, i.e.,

$$p^{\text{th}} = \frac{r+1}{2r}\left(\sqrt{1 + \frac{8r(r-1)}{(r+1)^2}} - 1\right), \qquad (9)$$

then the segment $0 \leqslant p_2 \leqslant p^{\text{th}}$ is an invariant set; in other words, there the flow is colinear to the edge. Clearly, $dp_2 > 0$ at $p_2 = 0$, while at $p_2 = p^{\text{th}}$ (where $f_2^D = f_1^C = f_3^C$) intra-group imitation leads to $dp_2 < 0$. Thus, there is a fixed point $0 < \hat{p} < p^{\text{th}}$ inside the segment, where the nullcline surface $dp_2 = 0$ intersects the edge $(1, p_2, 1)$.

To proceed in the search for fixed points located at the phase space boundary, we have finally to consider the three faces adjacent to the fully cooperative corner $(1,1,1)$, which form a class of equivalent faces. To allow for analytics, we explicitly consider one of these, say, the face defined by $p_3 = 1$, and, due to the symmetry by interchange of labels 1 and 2, we focus attention onto the triangular simplex $p_2 \leqslant p_1$, as in Sec. III A.

Note that we have already inferred the existence of a fixed point $(1, \hat{p}, 1)$ located at the edge. The replicator dynamics defines the flow Eq. (6) on it with the following transition probabilities (where τ is the characteristic time scale, here $\tau = 3\Delta f_{\text{max}}$):

$$P_3^+ = 0,$$

$$P_3^- = \frac{1}{\tau}\left[\sum_{i=1,2}(1-p_i)\left(f_i^D - f_3^C\right)\theta\left(f_i^D - f_3^C\right)\right],$$

$$P_1^+ = \frac{1-p_1}{\tau}\left(f_3^C - f_1^D\right)\theta\left(f_3^C - f_1^D\right),$$

$$P_1^- = \frac{p_1}{\tau}\left[(1-p_1) + (1-p_2)\left(f_2^D - f_1^C\right)\theta\left(f_2^D - f_1^C\right)\right],$$

$$P_2^+ = \frac{1-p_2}{\tau}\left[\left(f_3^C - f_2^D\right)\theta\left(f_3^C - f_2^D\right) + p_1\left(f_1^C - f_2^D\right)\theta\left(f_1^C - f_2^D\right)\right],$$

$$P_2^- = \frac{p_2}{\tau}\left[(1-p_2) + (1-p_1)\left(f_1^D - f_2^C\right)\theta\left(f_1^D - f_2^C\right)\right]. \qquad (10)$$

First we determine the region defined by $P_3^- = 0$ (equivalently, $f_1^D < f_3^C$), where the flow remains on the face. The condition $f_1^D = f_3^C$ defines a line $\tilde{p}_1(p_2)$ which intersects the edge $(p_1, 0, 1)$ at $\tilde{p}_1(0)$, and the symmetry line $(p, p, 1)$ at $\tilde{p}_1(p_2 = p_1)$. Note that the condition $f_2^D < f_3^C$ is also satisfied due to our restriction to the $p_2 \leqslant p_1$ triangle. The exact

analytical expression $\tilde{p}(q)$, for the borderline $\tilde{p}_1(p_2)$ is

$$\tilde{p}(q) = \frac{1 + rq}{2r}\left[\sqrt{1 + \frac{4r(r-1)(1+q)}{(1+qr)^2}} - 1\right], \quad (11)$$

from which the previous intersection ($\tilde{p}(0)$, $\tilde{p}(\tilde{p})$, and $p^{\text{th}} = \tilde{p}(1)$) points can be explicitly determined as functions of the model parameters.

Only to the left of this line the flow remains on the plane $p_3 = 1$. We then see that the invariant segment on the vertical edge, including the fixed point $(1, \hat{p}, 1)$, is disconnected from this region. On the contrary, the segment of the symmetry line below $\tilde{p}_1(p_2 = p_1)$ is an invariant set included in the region, where we now focus attention. At the lower bound $(0,0,1)$ of this segment the flow is positive, while at the upper one $(\tilde{p}_1, \tilde{p}_1, 1)$, cooperators in groups 1 and 2 imitate defectors, thus the flow is negative. Thus, there is a fixed point $(p^*, p^*, 1)$ inside this segment where the nullcline surface $dp_2 = 0$ intersects the symmetry line. Note that as this line invariant, the condition $dp_2 = 0$ entails that also $dp_1 = 0$. In fact, the intersection of the $dp_1 = 0$ nullcline with the face has two isolated points, namely, $(1,1,1)$ and $(p^*, p^*, 1)$, and the segment $(1, 0 \leqslant p_2 \leqslant p^{\text{th}}, 1)$, which is isolated from the region $P_3^- = 0$.

We have found, regarding stationary states, that besides the attractor $(0,0,0)$ and the repellor $(1,1,1)$, there are six saddle fixed points:

1. Three of them are located on the three cube edges adjacent to $(1,1,1)$, at $\hat{p}$; each of them has a stable manifold on the segment $[0, p^{\text{th}}]$ over the corresponding edge; it is clear that fluctuations along directions orthogonal to the edge are repelled away the fixed point, so that its unstable manifold is two-dimensional.

2. The last three fixed points are located at the symmetry lines of the three faces adjacent to $(1,1,1)$. The stable manifold for each of them is a two-dimensional (see top panel of Fig. 3 compact piece of the corresponding face. A fluctuation orthogonal to the face flows away, along the one-dimensional unstable manifold.

The three-dimensional visualization of the flow in the 3D phase space is dominated by the contraction of the interior phase space toward the fully defective state. However, located on three of the faces, there are codimension 1 invariant sets that are the stable manifolds of stationary states where one of the groups is fully cooperator, and the other two keep the same mixed state of strategic population. Also, located on the three cube edges, there are codimension 2 invariant segments that are the stable manifolds of stationary states with two full-C groups. We have arrived to these results through the use of exact and generalizable arguments.

C. Is more (groups) different?

The characterization of the deterministic trajectories in the phase space of our Markov model carried out for $m = 2$ and 3 in the previous subsections was obtained through arguments that are easily seen to hold for general values of $m \geqslant 2$, provided the m groups are equally sized, and thus the symmetry by label interchange is preserved. Thus, the following educated conjecture can be safely put forward:
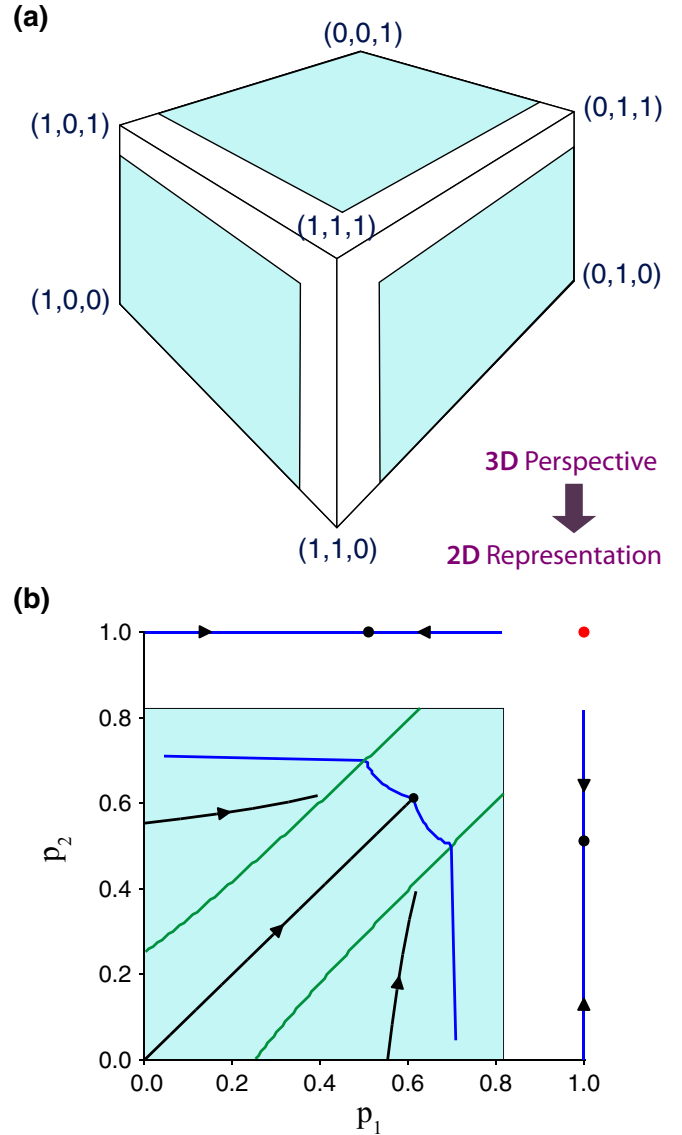

(a)

(b)

FIG. 3. Three-groups case. (a) Representation of the unit cube (p_1, p_2, p_3). Colored areas correspond to the planar regions, codimension 1 stable manifolds, where one of the three groups is fully cooperator. (b) Phase portrait for the three-groups case restricted to the plane $p_3 = 1$. The white area is the basin of the full defection state, while blue area corresponds to the stable manifold. The green lines correspond to the $f_1^C = f_2^D$ and $f_1^D = f_2^C$ singularities, blue lines to the nullclines, and arrows to trajectories. Inner black dot corresponds to the fixed point $(p^*, p^*, 1)$, and black dots located on the coordinate axes correspond to the fixed points $(\hat{p}, 1, 1)$, $(1, \hat{p}, 1)$. In this plot, $r = 4$. See the text for further details.

C1 Let n_f be any integer such that $1 \leqslant n_f \leqslant m - 1$, n_f representing the number of groups that are full-C. For any value of n_f there are a finite number saddle fixed points given by the combinatorial $C_{n_f}^m$, where n_f groups are full-C, and the rest $(m - n_f)$ groups are mixed groups with a fraction of cooperators $p = \hat{p}(n_f, m)$. These states are equivalent under label interchange. Any of them has a codimension n_f stable manifold, which corresponds to dimension $m - n_f$.

Note that the value of $\hat{p}(n_f = 1, m = 2)$ was computed in Sec. III A as Eq. (8), while in Sec. III B, we denoted by $\hat{p}$ what

in this general notation is $\hat{p}(n_f = 2, m = 3)$, and by p^* what is now termed $\hat{p}(n_f = 1, m = 3)$.

The replicator dynamics is given by

$$P_i^+ = \sum_{j=1}^{n_f} \frac{(1-p_i)(f_j^C - f_i^D)\theta[(f_j^C - f_i^D)]}{m\Delta f_{\max}},$$

$$P_i^- = \sum_{j=n_f}^{m} \frac{p_i(1-p_j)(f_j^D - f_i^C)\theta[(f_j^D - f_i^C)]}{m\Delta f_{\max}}. \qquad (12)$$

To obtain the value of $\hat{p}(n_f, m)$, we can express the fixed point condition as

$$\sum_{i=1}^{m} P_i^+ = \sum_{i=1}^{m} P_i^-. \qquad (13)$$

The function $\hat{p}(n_f, m)$ can be obtained, by substituting Eqs. (2) and (12) into the fixed point condition Eq. (13), as the positive solution (provided it is less than 1) to the quadratic equation,

$$ap^2 + bp + c = 0, \qquad (14)$$

with coefficients

$$a = (m - n_f)\left(\frac{m}{n_f} + r - 1\right),$$
$$b = m - (r-1)(m - 2n_f),$$
$$c = -(r-1)n_f.$$

One can quickly check that this result reproduces Eq. (8) for $m = 2$ and $n_f = 1$. No cooperator in the full-C groups will become a defector as long as its payoff is higher than the payoff of a defector in one of the "mixed-state" groups. From this consideration, one easily finds a lower bound $p^{th}(m, n_f)$ for the "fluctuation size" of the full-C groups. The value of this lower bound, below which stability is ensured, is the positive root of the quadratic equation,

$$a'p^2 + b'p + c' = 0, \qquad (15)$$

with coefficients

$$a' = (m - n_f)r,$$
$$b' = (2n_f - m)r + (m - n_f),$$
$$c' = -(r-1)n_f.$$

This provides the exact functional dependence of our lower bound of instability threshold p^{th} on all model parameters: m, n_f, and r. In the same way, Eq. (14) provides the exact functional dependence of the mixed cooperation level $\hat{p}$ on these parameters.

The mechanism that keeps a rather high average level of the groups is the positive contribution to P_j^+, for mixed groups j, from the imitation of full-C group members. It follows that there has to exist, at least, a full-C group for cooperation to persist. Therefore, this mechanism is fragile, for there is an instability threshold for fluctuations of the mixed groups cooperation, quantified by p^{th}, which restricts the stable manifolds of the fixed points to compact subsets on $(m - n_f)$-dimensional hypercubes. We will pay due attention to this fragility in Sec. IV.
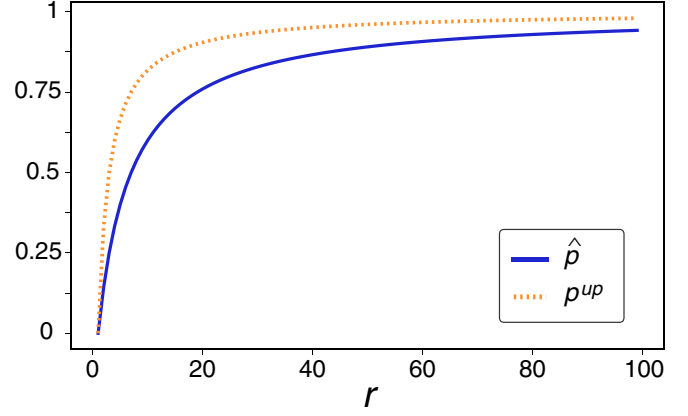

FIG. 4. Ground and upper fixed points. The graphs show the theoretical values of $\hat{p}$ (solid blue line) and p^{up} (dashed orange line) as a function of r. The number of groups has been fixed to $m = 6$, and the number of full-C groups to $n_f = 1$. See the text for further details.

In the states that the conjecture **C1** refers to, there are groups with two different values for the fraction of cooperators, and one wonders if more than two values for p_g are allowed in a fixed point. We now provide an argument supporting that "*more* (groups) *is different*" [28], regarding the fixed points of the Markov dynamics. More precisely, we will argue below that

C2 There are metastable fixed points where groups with more than two different values of group cooperation p_g coexist.

Consider one of the (type **C1**) stationary states with n_f groups at the level $p = 1$ and $(m - n_f)$ groups at the ground level $p = \hat{p}$. Now, choose one (the focal group, now on) of these latter groups, and change its fraction of cooperators to a value $p^{up} > \hat{p}$, such that:

1. $p^{up} < p^{th}$. This condition ensures that this change has no influence on the n_f groups in the full c level.

2. p^{up} is high enough to make impossible the imitation by cooperators in the focal group of defectors in ground groups.

3. The number m of groups is large enough to render very small the effects of the focal group on the equilibrium value of the ground level.

From the last two assumptions, one determines (see Appendix B) an analytical expression for p^{up} as a function of $\hat{p}$ and model parameters. Provided this value satisfies the first item above, this state can be taken as initial condition for the numerical direct integration of the Markov dynamics as a check of our third assumption, that hopefully refine both, p^{up} and $\hat{p}$, values. The results of the numerical direct integration are shown in Fig. 4.

The observations (that are quite generic, regarding variation of parameters) fully support the existence of fixed points with three levels of group cooperation, as well as its metastable character (see Sec. IV).

For a large number m of groups, there is no apparent reason that can forbid the existence of metastable fixed points with $\nu > 3$ levels of cooperation, at least for some range of model parameters and occupancies of the ν levels, provided the full cooperation level is not empty. Whenever two given levels are

known, the condition that cooperators in the upper level have a higher payoff than free-riders in the lower level groups fixes a threshold value, below which a new intermediate level of cooperation can be proved for metastability.

We see how, in this model case, "*more*" groups gives new kinds of metastable fixed points, by further breaking the permutation symmetry, and then "*is different*" [28]. We should at this point emphasize that all the fixed points we have found along this Sec. III, other than "all groups are full-C or full-D," are only invariant under a proper subset of group permutations. In other words, all of them are "symmetry-breaking" states. Also, all of them have at least one group of full cooperation.

IV. AGENT-BASED STOCHASTIC SIMULATIONS: FINITE GROUP SIZE FLUCTUATIONS AND PARAMETRIC SENSITIVITY

In the previous sections we have analyzed the dynamics of the Markov model. This is different from, although intended to mimic, the stochastic dynamics of agents that was a part of the model formal definition in Sec. II. Indeed, the Markov dynamics is the mesoscopic description. The m fractions p_g are group variables, while any microstate of the system of agents is really specified by $m \times n$ binary values (C or D). The number μ of agent microstates that are associated to a point $\{p_g\}$ ($g = 1, \cdots, m$) in the Markov phase space is

$$\mu(\{p_g\}) = \Pi_{g=1}^{m} \binom{n}{n p_g},$$

defining a nonuniform measure on the m-dimensional Markov phase space, which is highly concentrated at intermediate values of the group cooperation. When a stochastic evolution from an initial agent microstate is visualized as a trajectory on the mesoscopic phase space, large amounts of information are lost.

Also, while the Markov dynamics is deterministic, the updating of agents strategy is stochastic. Boundaries in Markov phase space that keep deterministic trajectories inside invariant regions are easily crossed by stochastic trajectories.

A convenient representation for trajectories of the many groups system is simply provided by the m graphs for $p_g(t)$ ($g = 1, \cdots, m$). In Fig. 5 we show the time evolution of the fraction of cooperators into the groups for two representative realizations corresponding to different group sizes, $n = 200$ (top) and $n = 1000$ (bottom), together with the theoretical values of $\hat{p}$ and p^{up}. As shown, the level of cooperation in the groups oscillates around points $\hat{p}$ and p^{up}. This fact is clearly displayed in Fig. 6, which represents the histogram of the fraction of cooperative agents into the different groups, after the transient period and accumulated over 100 independent realizations, for the same values than those used in the upper panel of Fig. 5.

The synergy factor r allows the agents to get a higher payoff for the same contribution, and helps enhancing the cooperators over the defectors. By increasing r, a given strategy yields a higher payoff for the same cooperation frequency, eventually allowing a cooperator in a group with cooperative population $n_C + x$ to beat a defector in a group with cooperative population n_C; it substantially helps the cooperation to
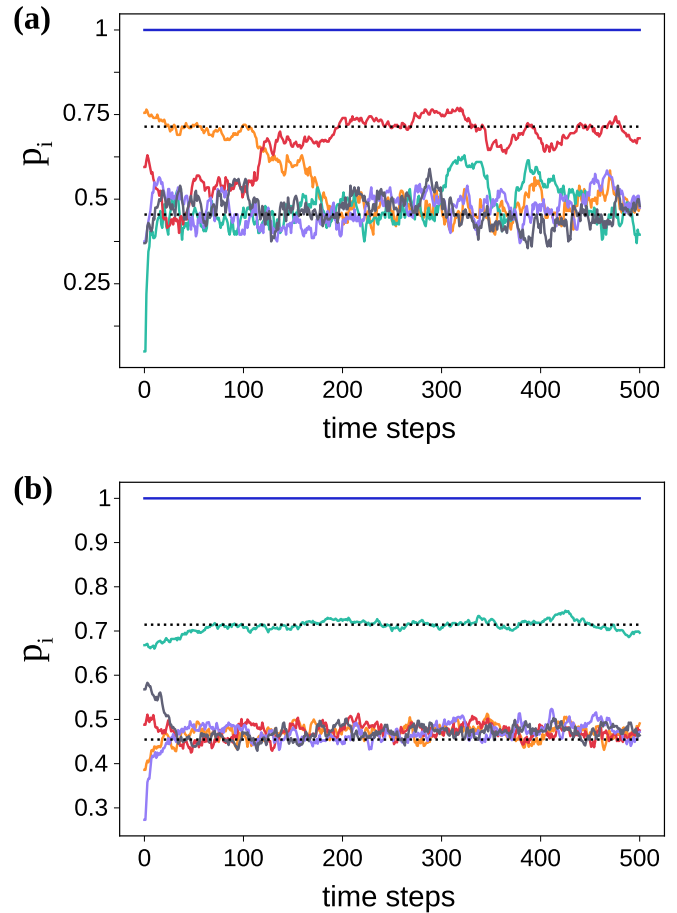

(a)

(b)

FIG. 5. Stochastic evolution of the system. The graphs show the time evolution of the fractions of cooperators for two representative realizations of the six-groups case, each solid line corresponding to a group. Dotted lines correspond to the theoretical predictions for $\hat{p}$ and p^{up}. (a) For small group sizes ($N = 200$) fluctuations allow groups to exchange their levels of cooperation. (b) For larger group sizes ($N = 1000$) one group clearly detaches from the rest to occupy the upper fixed point, and the fluctuations are not large enough to allow exchange. In these plots, $r = 6$.

maintain itself in a many-groups setup. In general, a full-C group can resist invasion by defectors if $r > n$. In that case, a single defector in a group of size n, which would earn the highest payoff a defector could expect, would still receive a payoff lower than cooperator in a full-C cooperator. For $r > n$, a group reaching full cooperation will not leave this state, as long as the fluctuations are not large enough to allow the system to escape from that state.

It is interesting to look at the fluctuations to infer whether cooperation is sustainable or not in this setup. As expected and shown in Fig. 5, fluctuations decay as the population size increases. For high enough values of n, the system is stable; the fluctuations around the fixed point do not allow a defector to beat a cooperator in the fully cooperative group, and cooperation is maintained. This remained true even for values of r of the order of n: Indeed, the higher the synergy factor, the less likely it becomes for a full-C cooperator to turn into a defector. We argue, by contrast, that finite-size effects can be dramatic: In some cases (typically $n \approx 10$),
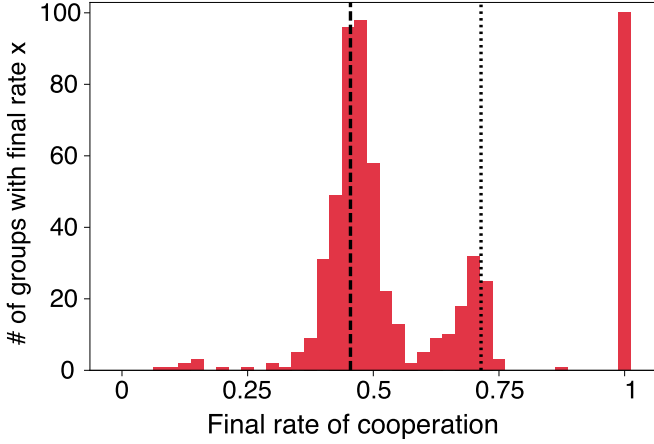
FIG. 6. Histogram of cooperation level in groups. The histogram counts the number of groups having a certain cooperation rate after 200 time steps. The counts have been accumulated over 100 independent realizations. Dashed (respectively, dotted) line corresponds to the theoretical predictions for $\hat{p}$ (p^{up}). Here, the parameters are assigned the same values as in the upper panel of Fig. 5: $m = 6$, $r = 6$, and $N = 200$. The initial cooperation rates in the non full-C groups are given by a Gaussian determined by $\mu = \sigma^2 = n/4$.

the fluctuations around the fixed point might allow a group to reach and overcome the critical population level, allowing invasion of the full-C group by defectors; the system cannot be considered stable anymore in this case. Actually, the fluctuations follow the inverse square root of the population size (from simulations $1/2\sqrt{n}$), which gives us an estimate of the robustness of the system as a function of the size of the groups.

V. DISCUSSION AND SUMMARY

In this study, we have taken a public goods approach to understand cooperative behavior in group-structured populations. In the proposed model, individuals are located in groups. The fitness of the individuals is related to their group performance, whereas the imitation takes place globally, that is, agents have a global vision and can imitate the most successful behaviors. We have shown that cooperation is maintained by a homogeneous group of cooperators. Note that this fully cooperative group does not necessarily have to be related to a physical group, but a social norm. In this sense, an interpretation of that group lies in the idea that cooperation in human societies is mainly based on social norms, and therefore the sustainability of the commons will be driven by these norms. For example, the legal enforcement of rules and the subsequent punishments, usually lie on a consensus about their legitimacy, that is, on the fact that those rules are backed by social norms [29]. The existence of a cooperative social norm [30] can be interpreted as a nonnull probability of cooperation, which in the proposed model is mathematically equivalent to a homogeneous group of cooperators. This social norm can be based either on moral principles or driven by both empirical and normative expectations [31–33].

The model is analytically solved through a Markovian approach, showing the existence of inner equilibria, invariant manifolds and thresholds for metastability. It is worth noting

that although both individual and group levels are based on the well-mixed approach, the model exhibits intermediate rates of cooperation under the replicator dynamics. Furthermore, agent-based stochastic simulations show group cooperation values fluctuating around the fixed points predicted by the Markov dynamics, numerically validating the analytical predictions. Finally, this study has implications on the effect of social norms and group selection on the sustainability of the commons.

APPENDIX A: JACOBIAN FOR THE TWO-GROUPS CASE

In this Appendix we present the general expression of the four terms of the Jacobian and the computation for J_{22} evaluated around the fixed point. Let us set $F_1 = f_2^D - f_1^C$ and $F_2 = f_1^D - f_2^C$. From Eqs. (5) and (6), it follows:

$$J_{11} = p_1 - (1 - p_2)\left[F_1 + p_1\left(\frac{\partial(F_1)}{\partial p_1}\right)\right]\theta(F_1),$$

$$J_{12} = p_1\left[(1 - p_2)\frac{\partial(F_1)}{\partial p_2} - F_1\right]\theta(F_1),$$

$$J_{21} = (1 - p_2)\left[F_1 + p_1\frac{\partial(F_1)}{\partial p_1}\right]\theta(-F_1)$$
$$- F_2 + (1 - p_1)\frac{\partial(F_2)}{\partial p_1},$$

$$J_{22} = p_1\left[(1 - p_2)\frac{\partial F_1}{\partial p_2} - F_1\right]\theta(-F_1) - (1 - p_1)F_2$$
$$- (1 - p_2) - p_2\left[(1 - p_1)\frac{\partial F_2}{\partial p_2}\right].$$

J_{22} value can be evaluated around the fixed point p^* computed in Eq. (8):

$$J_{22} = 2r\left(\frac{r + 1}{r - 1} - 1\right),$$

which is always negative for $r \geqslant 0$.

APPENDIX B: ESTIMATION OF THE UPPER FIXED POINT

In this Appendix we estimate the value of the metastable fixed point p^{up} discussed in point **C2** of Sec. III C. Let us consider more than two groups $m > 2$ and, at least, one full-C group and more than one mixed groups $1 \leqslant n_f \leqslant m - 1$. The polynomial equation for the population in the upper fixed point p^{up} is found by imposing **C2** conditions in the replicator

dynamics, which drastically reduces the P^- term and thus results in a higher value for the cooperation frequency:

$$e(p^{\text{up}})^2 + gp^{\text{up}} + h = 0, \tag{B1}$$

with

$$e = n^{\text{up}}\left(\frac{n^{\text{up}}}{n_f} + r\right),$$

$$g = \alpha(n_f + \hat{n}\hat{p}) - (r-1)n^{\text{up}},$$

$$h = (1-r)(s + \hat{n}\hat{p}).$$

where $\hat{n}$ and $\hat{p}$ represent the number of groups in the lowest fixed point and their cooperation frequency, respectively, and n^{up} and p^{up} represent the number of groups in the upper fixed point and their cooperation frequency. After a first estimation of those values based on the assumption that the variation of the ground value $\hat{p}$ is small, a better characterization of the fixed points can be obtained by refining the values of the fixed points and of the mean cooperation iteratively until convergence. The final value of the mean cooperation in mixed groups $\bar{p}$ is given by

$$\bar{p} = \frac{n^{\text{up}}p^{\text{up}} + (m - n_f - n^{\text{up}})\hat{p}}{m - n_f}. \tag{B2}$$

Since it is possible to compute the value of the mean cooperation frequency in the general case, we can compute the upper limit p^{th} allowed for sustainability of cooperation. Let f_C^f be the fitness of a cooperator in the full-C group, and f_D^{th} be the fitness of a defector in the pth group. From the limit condition,

$$f_C^f = f_D^{\text{th}}, \tag{B3}$$

it follows

$$p^{\text{th}}r = (r-1), \quad p^{\text{th}} = 1 - \frac{1}{r}. \tag{B4}$$

If in a group the cooperation frequency overcomes this value, then cooperators in the fully cooperative group can turn into defectors and the system would be in the basin of attraction of the fully defective state.

APPENDIX C: UNEQUAL-SIZED GROUPS

An interesting way toward the generalization of this model is to introduce disparity in groups sizes. Note that, considering that all the agents have the same probability to be chosen for imitation, larger groups will be more influential than smaller ones. This asymmetry may allow defectors in large groups to invade small cooperative groups.

As in the case of equal group sizes, without a fully cooperative group, the system is in the basin of attraction of the full-defection state.

Since the imitation probabilities depend on the size of each group, the probabilities for a group i to increase and decrease its fraction of cooperators p_i by $1/n_i$ become, respectively,

$$P_i^+ = \frac{n_i(1-p_i)}{\tau} \sum_{\substack{j=1 \\ j \neq i}}^m n_j p_j \left(f_j^C - f_i^D\right) \theta\left(f_j^C - f_i^D\right),$$

$$P_i^- = \frac{n_i p_i}{\tau} \sum_{j=1}^m n_j p_j \left(f_j^D - f_i^C\right) \theta\left(f_j^D - f_i^C\right), \tag{C1}$$

where n_i represents the total number of agents of group i, N represents the total number of agents in the whole system, and τ is the characteristic time scale involving Δf_{max} and N.

The expression for the fixed point p^* corresponding to all the mixed groups (non full-C groups) sharing the same value p_i is obtained by setting $P_i^+ = P_i^-$:

$$(N - n_f)^2 p^*(1 - p^*) = n_f(N - n_f)(1 - p^*)(r - p^*r), \tag{C2}$$

which yields the value of the fixed point:

$$p^* = \frac{n_f(r-1)}{N - n_f}. \tag{C3}$$

A full-C group will resist invasion by defectors if

$$f_C^f > f_D^g \quad \Rightarrow \quad \frac{r-1}{r} > p_g^* n_g, \tag{C4}$$

where n_g represents the size of the biggest group, excluding the full-C group.

Condition Eq. (C4) relates the size of the fully cooperative group, the size of the biggest mixed group and the synergy factor r. It expresses whereas the full-C group can survive, and therefore, whether cooperation is stable in such a system or not.

[1] J. M. Smith *et al.*, *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, UK, 1982).

[2] E. Pennisi, Science **309**, 93 (2005).

[3] J. S. Chuang, O. Rivoire, and S. Leibler, Science **323**, 272 (2009).

[4] J. Gore, H. Youk, and A. Van Oudenaarden, Nature **459**, 253 (2009).

[5] J. H. Kagel and A. E. Roth, *The Handbook of Experimental Economics, Vol. 2: The Handbook of Experimental Economics* (Princeton University Press, Princeton, NJ, 2016).

[6] P. Kollock, Ann. Rev. Sociol. **24**, 183 (1998).

[7] R. L. Trivers, Q. Rev. Biol. **46**, 35 (1971).

[8] J. H. Fowler, Nature **437**, E8 (2005).

[9] M. A. Nowak and K. Sigmund, Nature **393**, 573 (1998).

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

[11] W. D. Hamilton, J. Theor. Biol. **7**, 17 (1964).

[12] V. C. Wynne-Edwards, Technical Report No. QL752 W9 1962a (University Aberdeen, Aberdeen, Scotland, UK, 1962).

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

[14] D. S. Wilson and E. Sober, Behav. Brain Sci. **17**, 585 (1994).

[15] A. Rapoport, A. M. Chammah, and C. J. Orwant, *Prisoner's Dilemma: A Study in Conflict and Cooperation*, Vol. 165 (University of Michigan Press, Ann Arbor, MI, 1965).

[16] M. Doebeli and C. Hauert, Ecol. Lett. **8**, 748 (2005).

[17] J. Andreoni and J. H. Miller, Econ. J. **103**, 570 (1993).

[18] R. Cooper, D. V. DeJong, R. Forsythe, and T. W. Ross, Games Econ. Behav. **12**, 187 (1996).

[19] C. Gracia-Lázaro, A. Ferrer, G. Ruiz, A. Tarancón, J. A. Cuesta, A. Sánchez, and Y. Moreno, Proc. Natl. Acad. Sci. USA **109**, 12922 (2012).

[20] M. Perc, J. Gómez-Gardeñes, A. Szolnoki, L. M. Floría, and Y. Moreno, J. R. Soc. Interface **10**, 20120997 (2013).

[21] U. Motro, J. Theor. Biol. **151**, 145 (1991).

[22] F. C. Santos, M. D. Santos, and J. M. Pacheco, Nature **454**, 213 (2008).

[23] A. Szolnoki and M. Perc, Europhys. Lett. **92**, 38003 (2010).

[24] C. Gracia-Lazaro, J. Gomez-Gardenes, L. M. Floria, and Y. Moreno, Phys. Rev. E **90**, 042808 (2014).

[25] H. Gintis, *Game Theory Evolving: A Problem-centered Introduction to Modeling Strategic Behavior* (Princeton University Press, Princeton, NJ, 2000).

[26] F. C. Santos and J. M. Pacheco, Phys. Rev. Lett. **95**, 098104 (2005).

[27] J. Hofbauer and K. Sigmund, Bull. Amer. Math. Soc. **40**, 479 (2003).

[28] P. W. Anderson, Science **177**, 393 (1972).

[29] E. Fehr and U. Fischbacher, Trends Cognit. Sci. **8**, 185 (2004).

[30] J. M. Weber and J. K. Murnighan, J. Personal. Soc. Psychol. **95**, 1340 (2008).

[31] W. Güth and H. Kliemt, Metroeconomica **45**, 155 (1994).

[32] C. Bicchieri, *The Grammar of Society: The Nature and Dynamics of Social Norms* (Cambridge University Press, Cambridge, UK, 2005).

[33] M. Chudek and J. Henrich, Trends Cognit. Sci. **15**, 218 (2011).