J. theor. Biol. (2000) **206,** 465–485
doi:10.1006/jtbi.2000.2130, available online at http://www.idealibrary.com on IDEAL®

On the Dynamic Persistence of Cooperation: How Lower Individual Fitness Induces Higher Survivability

GUY SELLA*†‡ AND MICHAEL LACHMANN§

*School of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel, †Department of Biological Sciences, Stanford University, Stanford, CA 94305, U.S.A., §Santa Fe Institute, Santa Fe, NM 87501, U.S.A.

(*Received on* 16 *March* 1999, *Accepted in revised form* 20 *June* 2000)

We study a model in which cooperation and defection coexist in a dynamical steady state. In our model, subpopulations of cooperators and defectors inhabit sites on a lattice. The interactions among the individuals at a site, in the form of a prisoner's dilemma (PD) game, determine their fitnesses. The chosen PD payoff allows cooperators, but not defectors, to maintain a homogeneous population. Individuals mutate between types and migrate to neighboring sites with low probabilities. We consider both density-dependent and density-independent versions of the model. The persistence of cooperation in this model can be explained in terms of the life cycle of a population at a site. This life cycle starts when one cooperator establishes a population. Then defectors invade and eventually take over, resulting finally in the death of the population. During this life cycle, single cooperators migrate to empty neighboring sites to found new cooperator populations. The system can reach a steady state where cooperation prevails if the global "birth" rate of populations is equal to their global "death" rate. The dynamic persistence of cooperation ranges over a large section of the model's parameter space. We compare these dynamics to those from other models for the persistence of altruism and to predator–prey models.

1. Introduction

Explaining the evolution and persistence of cooperation is a central problem in evolutionary biology and the social sciences (Axelrod, 1984). A cooperating individual has higher fitness within a group of cooperators than it has in isolation. Nevertheless, cooperative behavior often entails a fitness cost. Consequently, a defecting individual—one that enjoys the cooperation of others but abstains from cooperative behavior—will have an immediate selective advantage over cooperators. This advantage renders a population of cooperators susceptible to invasion and take-over by defecting individuals. Therefore, the persistence and evolution of cooperation seems to face an intrinsic instability.

The interaction between cooperators and defectors is often formalized in terms of the game known as the prisoners' dilemma (PD) (Weibull, 1995). The PD payoff matrix appears in Table 1. This is a symmetric game between two players, where each player has two possible strategies: defect or cooperate. The game is set up such that, for any strategy of the opponent, a defector has a greater payoff than a cooperator, but if both players cooperate they have a greater payoff than if both defect. In the framework of evolutionary game theory (Weibull, 1995) this game can be used in order to study the evolution and

‡Author to whom correspondence should be addressed.

0022–5193/00/200465 + 21 $35.00/0

TABLE 1

The prisoner's dilemma payoff matrix: Each box describes the payoff for a possible two-player interaction. The left entry refers to the player employing the strategy above, while the second refers to the player employing the strategy listed on the side. The payoffs are set such that $\delta > \alpha > \beta > \gamma$. Under this condition, the best strategy is to defect independent of the other player's strategy. If both players defect, however, they both receive lower payoffs than if they both cooperate

	c	d
c	(α, α)	(δ, γ)
d	(γ, δ)	(β, β)

persistence of cooperation. One can assume that individuals in a population of cooperators and defectors interact randomly in pairs, and that the fitness of an individual is determined by the payoffs it receives in its interactions. The population dynamics under these assumptions have two principal characteristics:

1. A population consisting only of cooperators grows faster than a population consisting only of defectors (this follows from the relation $\alpha > \beta$ in the payoff matrix).
2. In any population with both types, a defector has a higher fitness than a cooperator (this follows from the relations $\delta > \alpha$ and $\beta > \gamma$).

How can cooperation evolve and persist if defection is always the locally favored strategy? We mention three main categories of answers to this question, though the distinction between them is not always sharp. The first category is the individual centered approach, where cooperation persists because it eventually confers a fitness advantage at the level of the individual. Models incorporating reciprocity (Axelrod, 1984), partnership (Cooper & Wallace, 1998), or the handicap principle (Roberts, 1998) fall into this class. The second category is kin selection (Hamilton, 1964). This includes models of kin recognition (Axelrod, 1984), models where kin interaction results from individual behavior in a spatial context, and more generally models of statistical kinship (Eshel & Cavalli-Sforza, 1982). The third category consists of structured population dynamic models. It includes among others, the hay stack model (Maynard-Smith, 1964; Wilson, 1987), models of the founder effect (Cohen & Eshel, 1976), and models of the neighbor effect (Eshel, 1971). Nowak & May (1992) introduced a family of models which combines aspects of all three categories mentioned above, and was studied intensively during the last decade. In these models (Nakamaru & Iwasa, 1997, 1998; Ferriere & Michod, 1995; Oliphant, 1999), individuals occupy lattice sites and play the iterated or non-iterated PD game with their neighbors on the lattice. The model we present here falls into the third category. The dynamics which maintain cooperation in our model, which we describe below, distinguish it from earlier models in this category.

In a review about group selection (Maynard-Smith, 1976), Maynard Smith mentions a predator–prey model (Maynard-Smith, 1974) and claims that it is analogous to a model for the persistence of altruism (or cooperation in our terminology). In that model, isolated patches may be in one of the following three states: E—empty, containing neither prey nor predator, H—containing prey only, and M—containing both prey and predator. An empty patch may be colonized by prey that migrate from a different patch, thus changing its state from E to H. A patch in state H may be colonized by migrating predators, changing its state from H to M. In a patch in state M, the predators eventually exhaust the prey and die, thus changing the state of the system from M to E. Having studied such systems with computer simulations, Maynard Smith concludes that: "… such models can rather easily give persistent coexistence of predator and prey; that is, persistence does not require a particularly careful choice of parameters". In such a state of persistence, each patch goes through a series of transitions $E \to H \to M \to E \to \cdots$ indefinitely. In this paper, we show that the underlying dynamics of interaction between cooperators and defectors can lead to the persistence of cooperation in a dynamic mode similar to that described by Maynard Smith.

In the model presented in Section 2, subpopulations of cooperators and defectors cohabit

sites on a lattice. The random interaction among the individuals at a site determine their fitness, based on PD payoffs. The standard PD condition that $2\alpha > \gamma + \delta$, meaning that a group of cooperators has a higher average fitness than any other group, does not affect the dynamics, and therefore we do not impose it. Individuals mutate between types and migrate to a neighboring site with low probabilities. The fitness function at a site may also depend on the population density at the site. It is important to stress that in this model the population size at a site is finite and varies over time. The population at a given site may die out, thereby leaving the site empty until it is occupied again by a migrating individual.

We seek the conditions for the persistence of cooperation when subpopulations consisting solely of defectors are doomed to face extinction, while subpopulations consisting solely of cooperators are capable of persisting. In terms of the PD payoffs, this qualitative asymmetry occurs when $\alpha > 1 > \beta$. We show that under this assumption, the population dynamics at a site takes the form of a life cycle. The life cycle begins when a cooperator migrates into an empty site and founds a cooperator population. This population is later invaded by a defector, which is either a mutant or a migrant. Defectors then take over, after which the population dies and the life cycle ends. In Section 3, we study the life cycles that emerge in both the density-dependent and density-independent models. The life cycles are analysed in terms of two variables that are later used to characterize the conditions for the dynamic persistence of cooperation in the system: R—the ratio of the total number of cooperators to the total number of defectors over the duration of a life cycle; and M—the total number of individuals that migrate out of a site over the duration of a life cycle.

Over a large range of the parameter values, cooperation persists in a steady state where on average one cooperator migrates to an empty neighboring site during a single life cycle, thus initiating a new cycle before the original cycle ends. In such a steady state the global "birth" rate of cooperating populations balances their "death" rate due to defector take-over. In Section 4, we use computer simulations to study the regions of dynamical persistence in terms of the life cycle variables R and M. We also construct a simplified model for the density-dependent model which is similar to the predator–prey model described above. Using this simplified model, we construct a mean-field approximation in which we analytically derive the conditions for dynamical persistence, and a higher-order approximation incorporating spatial correlations. These approximations explain the shape of the boundary that separates the regions in which cooperation persists from the regions in which it does not in the R and M space. For the density-independent model, we show that increasing the payoff for the interaction between cooperators, α, can increase the number of cooperators in the system, the number of sites they occupy, and their number relative to defectors. Moreover, increasing this payoff beyond a critical value results in the extinction of cooperation. These results are explained both intuitively and on the basis of analytical derivations.

2. The Model

We introduce an evolutionary model of finite subpopulations inhabiting sites on an infinite two-dimensional rectangular lattice. Each individual in a subpopulation is either a cooperator or a defector, where these behaviors are genetically determined (variables referring to these will be marked with subscript c for cooperators and d for defectors). The dynamics proceed from selection with absolute fitnesses, mutation and diffusion.

The fitness of an individual stems from its interactions with other individuals at the same site on the lattice, according to the PD payoff matrix in Table 1. We assume a PD-type interaction, with the asymmetry described in the introduction. The payoffs must therefore satisfy the following conditions:

$$\delta_{(dc)} > \alpha_{(cc)} > 1 > \beta_{(dd)} > \gamma_{(cd)} \geqslant 0 \qquad (1)$$

(henceforth we omit the subscripts on these parameters). Consider a site with n_c cooperators and n_d defectors interacting at random. The absolute fitness functions of cooperators and defectors at time t, which measure the average growth in

a time step Δ, are†

$$f_c(t) = g(n(t))\left(\alpha \frac{n_c(t)}{n(t)} + \gamma \frac{n_d(t)}{n(t)}\right), \tag{2}$$

$$f_d(t) = g(n(t))\left(\delta \frac{n_c(t)}{n(t)} + \beta \frac{n_d(t)}{n(t)}\right), \tag{3}$$

where $n = n_c + n_d$ is the total population at the site, and $g(n)$ is a function reflecting the density dependence. We will assume that $g(n) \leqslant 1$, but that $\alpha g(n_c) > 1$ for a population size smaller than the carrying capacity $n^* > 0$ (the density-independent model corresponds to $g(n) = 1$).

These fitness functions are characterized by two principal features:

1. A homogeneous cooperator population grows faster than a homogeneous defector population. Moreover, a homogeneous cooperator population is capable of maintaining itself while a homogeneous defector population is not. These properties follow from the relations $f_d(0, n_d) = \beta g(n_d) < 1$ for any $n_d \neq 0$, and $f_c(n_c, 0) = \alpha g(n_c) > 1$ for n_c below the carrying capacity n^*.
2. A defector has a higher fitness than a cooperator in any population structure, since $\delta > \alpha$ and $\beta > \gamma$ imply $f_d(n_c, n_d) > f_c(n_c, n_d)$ for all n_c and n_d.

The population growth at a site is described by a stochastic process, resulting in the absolute fitness functions of eqns (2 and 3). The following analytic results depend only on the average growth rates, and not on the specific stochastic process that realizes them. In our simulation, we assume that an individual with fitness $f = n + x$, where n is a nonnegative integer and $0 < x < 1$, is represented in the next time step by $n + 1$ individuals with probability x, and by n individuals with probability $1 - x$.

† Note that the fitness functions include the interaction of an individual with itself. Not including an individual in the calculation of its fitness requires a separate definition of the fitness of an individual in isolation, but does not change qualitatively any of the results in this paper.

Mutation and migration are incorporated in each time step as follows:

1. The subpopulations at all sites grow stochastically according to the fitness functions.
2. Each individual may mutate to become the other type with probability $\mu \ll 1$.
3. Each individual may migrate to one of its neighboring sites with probability $D \ll 1$. We assume the von-Neumann neighborhood of four neighbors on a rectangular lattice.

The expected values for $n_c(t + \Delta)$ and $n_d(t + \Delta)$ at a site given $n_c(t)$ and $n_d(t)$ are then

$$\begin{aligned} &E(n_c(t+\Delta)|n_c(t), n_d(t)) \\ &\quad = (1-D)[(1-\mu)f_c(t)n_c(t) + \mu f_d(t)n_d(t) \\ &\qquad + D\left[\frac{1}{4}\sum_{n.n.} f_c^{n.n.}(t)n_c^{n.n.}(t) - f_c(t)n_c(t)\right] \\ &\qquad + O[(D+\mu)^2], \\ &E(n_d(t+\Delta)|n_c(t), n_d(t)) \\ &\quad = (1-D)[(1-\mu)f_d(t)n_d(t) + \mu f_c(t)n_c(t) \\ &\qquad + D\left[\frac{1}{4}\sum_{n.n.} f_d^{n.n.}(t)n_d^{n.n.}(t) - f_d(t)n_d(t)\right] \\ &\qquad + O[(D+\mu)^2], \end{aligned} \tag{4}$$

where superscript *n.n.* denotes the values at nearest-neighboring sites. We take $\Delta \ll 1$ where 1 denotes the duration of an average generation. This implies that the fitness coefficients α, β, γ, δ and the corresponding fitness functions are close to 1. Furthermore, it means that the stochastic process approaches a process continuous in time. When $\Delta \ll 1$, $\mu \ll 1$ and $D \ll 1$, however, our scheme is equivalent to any other reasonable scheme that incorporates selection, mutation and diffusion, and to other reasonable schemes for the stochastic selection.

3. Local Behavior—The Life Cycle

We begin the analysis by considering the local behavior at one site. This behavior can be described in terms of a typical *life cycle*. A life cycle for the density-independent model (d.i.) is described in Fig. 1. It begins when one cooperator migrates to an empty site. The population of cooperators then begins to grow. The growth rate depends on $f_c(n_c, 0) = \alpha g(n_c)$. At some time denoted by t_f, the first defector appears from mutation or migration and defectors begin to take over. At some stage when defectors dominate the population, the fitness of cooperators drops below 1 and their number starts decreasing. Some time afterwards, as the frequency of cooperators decreases the fitness of defectors approaches β and in the process becomes less than 1. The life cycle ends at time t_e when the last defector dies, some time after the last cooperator disappeared.

Every life cycle ends with the death of the whole population at a site after a finite time. Therefore, for cooperation to persist, new life cycles have to be founded at a rate that balances their termination. The conditions for the existence of such a steady state are studied in Section 4. These conditions will be stated in terms of parameters characterizing the life cycle. A natural choice of parameters relevant for the global dynamics is the number of cooperators and defectors that migrate out of a site during a life cycle; we denote them as M_c and M_d. In order to derive these parameters from a description of the life cycle, we define

$$S_c \equiv \int_0^{t_e} n_c(t)\,\mathrm{d}t, \tag{5}$$

$$S_d \equiv \int_0^{t_e} n_d(t)\,\mathrm{d}t. \tag{6}$$

M_c and M_d are then given by

$$M_c = DS_c, \tag{7}$$

$$M_d = DS_d. \tag{8}$$

Two equivalent parameters which turn out to be useful are

$$M \equiv M_c + M_d = D(S_c + S_d), \tag{9}$$

$$R \equiv \frac{M_c}{M_d} = \frac{DS_c}{DS_d} = \frac{S_c}{S_d}. \tag{10}$$

Here M corresponds to the total number of units migrating from a site during a life cycle and R corresponds to the ratio of cooperators to defectors in the life cycle. The dependence of the global dynamics and in particular of the persistence of cooperation on R and M, will be studied in Section 4. In this section, we study the population dynamics at a site, focusing on the factors determining R and M. We find that in the d.i. model R depends on α, β, γ and δ, and S_d depends mainly on μ and D; while in the density-dependent (d.d.) models R depends on μ and D, and S_d is essentially constant.

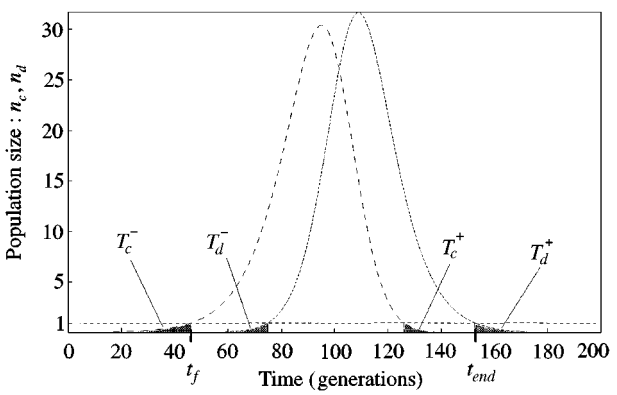

FIG. 1. A schematic representation of a life cycle: population size vs. time. Changing mutation rate scales the whole cycle, leaving the relations between the number of cooperators and defectors the same: (– – –) cooperators; (·····) defectors.

3.1. THE DENSITY-INDEPENDENT MODEL

In the d.i. model the dynamics [eqn (4)] are homogeneous to the first order in n_c and n_d. Thus, scaling n_c and n_d by a factor Λ at some time will just scale S_c and S_d by the same factor, leaving R unchanged. It is not hard to show, that taking two life cycles that vary only in the time in which the first defector invades, i.e. taking $\tau_f > t_f$, is

equivalent to scaling S_c and S_d by the factor†

$$\Lambda = \alpha^{[1 + (\log \alpha / \log (\alpha/\delta))](\tau_f - t_f)}. \quad (11)$$

The only significant effect (i.e. not $O[\mu, D]$) that mutation and migration have on the life cycle is in determining t_f. Therefore, we conclude that $R = R(\alpha, \beta, \gamma, \delta) + O[\mu, D]$. Result 1 provides an explicit expression for $R(\alpha, \beta, \gamma, \delta)$.

Result 1. *In the density-independent model the ratio R of the average number of migrating cooperators to the average number of migrating defectors in a life cycle is*

$$R(\alpha, \beta, \gamma, \delta, \mu, D) = \frac{E(M_c)}{E(M_d)} = \frac{DE(S_c)}{DE(S_d)}$$

$$= \frac{1-\beta}{\alpha - 1}\frac{\alpha - \gamma}{\delta - \beta} + O[\mu, D]. \quad (12)$$

We prove this in Appendix A.

A careful look at eqn (12) reveals that reducing α (but maintaining the condition $\alpha > 1$) while leaving every other parameter fixed can yield a larger ratio of cooperators to defectors, since

$$\left.\frac{\partial R}{\partial \alpha}\right|_{\beta, \gamma, \delta} < 0 \quad \text{for } \delta > \alpha > 1 > \beta > \gamma. \quad (13)$$

It seems reasonable and it will be shown later, that the larger R is, the more likely it is that cooperation could persist in the system. This hints at the possibility that in certain parameter regions of the d.i. model, decreasing α while leaving all the other parameters fixed will transform the global behavior from a state where cooperation cannot persist to a state where it can. Such behavior is seen in simulation results in Figs 3, 4, 5, 9, in Section 4. Thus, in this model lower cooperator fitness may induce higher survivability!

This effect would not have been anticipated from an individual centered perspective. On the other hand, when the life cycle is considered, there is a simple explanation for this effect: The integral number of defectors during the life cycle S_D strongly depends on the number of cooperators at the time the first defector invades $n_c(t_f)$. Therefore, cooperators can increase their fraction by maximizing their integral S_c while keeping $n_c(t_f)$ fixed. This explains why R increases when α is smaller. Roughly speaking, the moral is that when surrounded by defectors, keeping a low profile might be a good idea.

3.2. THE DENSITY-DEPENDENT MODEL

We consider an example of a d.d. model with

$$g(n) = \begin{cases} 1, & n < \dfrac{n_{max}}{\delta}, \\ \dfrac{n_{max}}{\delta n}, & n \geqslant \dfrac{n_{max}}{\delta}. \end{cases}$$

In this model, the population size is bounded by n_{max}.

The life cycle for this model is described in Fig. 2. Unlike the d.i. life cycle, in this case the number of cooperators stabilizes after a finite time $t(\alpha)$ on $c(\alpha)n_{max}$, where $c(\alpha) \equiv \alpha/\delta$. The shape

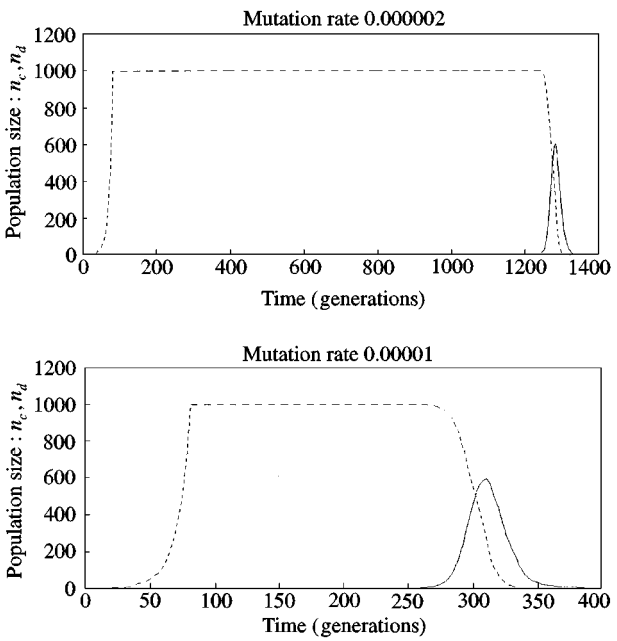

FIG. 2. The density-dependent life cycle at a site for the model in the example. The graph shows population size vs. time, for two different mutation rates $\mu = 0.00001$ and 0.000002: (——) defectors; (– – –) cooperators.

† In this argument we are ignoring the tails T_c and T_d shown in Fig. 1. These tails represent the parts of an extrapolated life cycle in which n_c and n_d drop below one. Changing t_f moves parts of these tails into S_c and S_d. However, this effect can be shown to change R only by $O[\mu, D]$.

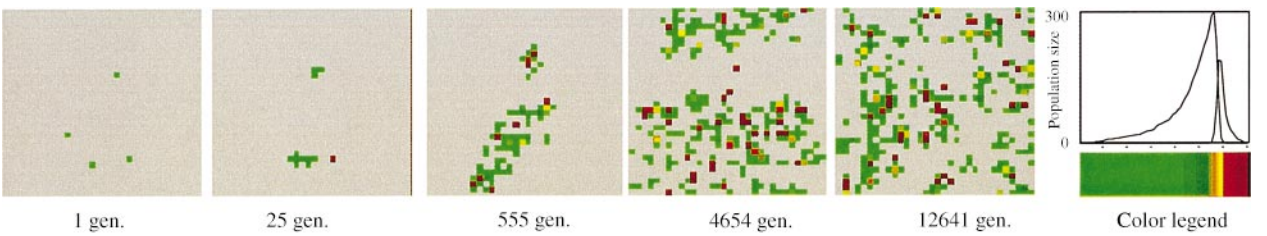

FIG. 3. Simulation of the density-independent model on a 32×32 lattice with periodic boundary conditions. The parameters for this simulation were: $\alpha = 1.01$, $\beta = 0.95$, $\gamma = 0.8$, $\delta = 1.2$, $\mu = 0.0001$ and $D = 0.0003$. The state of the sites of the lattice are given according to the color key on the right, which corresponds to the different stages in the life cycle for these parameter values. The simulation begins with single cooperators inhabiting a few sites. The system reaches a non-trivial steady state, where cooperation persists dynamically.

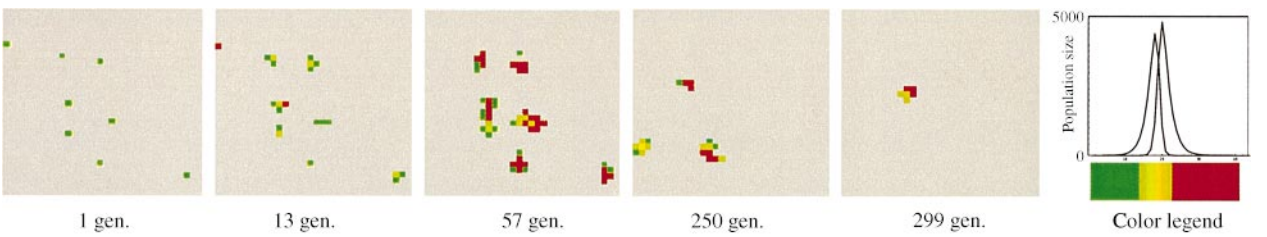

FIG. 4. In this simulation we took: $\alpha = 1.06$, where all the other parameters are identical to those in the simulation in Fig. 3. In this case, the system reaches the trivial—all empty steady state. This is a case where taking higher cooperator fitness results in the extinction of cooperation. This effect is discussed further in Section 4.3.

of this life cycle implies that as long as $t_f > t(\alpha)$, changing parameters so that t_f increases, will increase S_c, leaving S_d constant. This means that one could make R bigger than any $R^* > 0$ by picking a large enough t_f. Taking a large t_f simply means taking a small enough μ and D. Hence, for this example we conclude, that R can assume any large value and S_d remains essentially constant if μ and D are taken to be small enough.

This behavior is generic, it characterizes a class of d.d. models we define in Appendix B as models of type 1. In Appendix B, we apply the same reasoning used in the example to prove the following result:

Result 2. *Given a density-dependent model of type* 1 *and* $R^* \geqslant 0$, *taking* μ *and* D *such that*

$$D + \mu \leqslant \frac{c(\alpha)}{c(\alpha) n_{max} t(\alpha) + \widetilde{S_d}\, R^*}$$

ensures that R *is bounded from below by* R^*, *where* $\widetilde{S_d}$ *is* $E(S_d)$ *for the density-independent model with the same parameters* α, β, γ, δ, *and initial conditions* $\widetilde{n_c}(0) = n_{max}$, $\widetilde{n_d}(0) = 1$.

This means that for any such model and parameters α, β, γ, and δ, any desired ratio R may be attained by taking small enough μ and D.

4. Global Behavior

In the systems we studied, there are two types of steady states for the global behavior: the trivial steady state where all sites are empty, and non-trivial steady states in which cooperation persists globally. The behavior of two simulations of systems with d.i. dynamics is presented in Figs 3 and 4. In the first simulation, we start with a few sites inhabited by one cooperator each, and cooperation spreads to establish a non-trivial steady state. This steady state is dynamic in nature; cooperation persists even though each cooperator population eventually dies out. This can be seen in Fig. 5, where the population size at one site is described as a function of time. In the system described in Fig. 4, the population also starts with a few sites inhabited by one cooperator, but in this case, populations do not seed new ones at a rate that balances their rate of destruction by defectors from within and without. All the subpopulations in this system eventually die out, leaving it in the steady state where all the sites are empty.

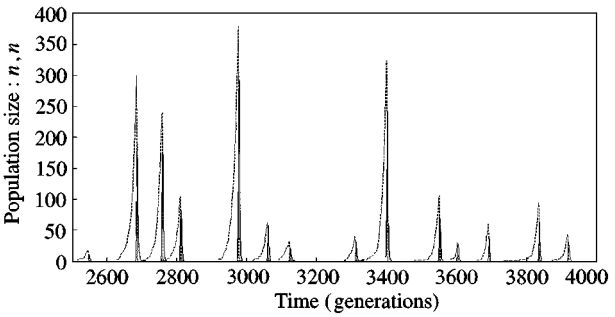

FIG. 5. The population at a specific site as a function of time, for the same simulation presented in Fig. 3. The life cycles vary in size due to the stochasticity in the t_f's. This stochasticity is caused by variations in the environment and in mutation. However, the shape of the different life cycles is similar, in correspondence with the scaling properties discussed in Section 3.1: (——) defector population; (·····) cooperator population.

A well-defined non-trivial steady state requires an infinite lattice. Yet our simulations occur on a finite lattice. We argue that when it exists, the non-trivial steady state is the only steady state, and thus it will be attained in any reasonable choice of initial conditions. We do not prove this claim on our system. We show, however, that the fixed point analog to the non-trivial steady state in the mean-field approximation to the d.d. dynamics is the only stable fixed point in the system when it exists. Due to the stochasticity, any finite realization of the model on a finite lattice will always end up in the state where all the sites are empty. Nevertheless, the non-trivial steady state has a pronounced signature in the finite realizations of the model. In the parameter ranges corresponding to the non-trivial steady state, the duration in which cooperation persists in the finite system grows very fast with the size of the lattice. In Appendix D, we describe the criteria we use to determine when the finite simulations reach a state corresponding to the steady state. In the parameter ranges where these criteria hold, the system never reaches the empty state, in thousands of simulations lasting hundreds of thousands of generations each. Consequently, we can study the regions in which cooperation persists using simulations on finite

systems, and assume that large systems have dynamical behavior that is independent of initial conditions.

In these models, the global dynamics derive from an interplay between the local dynamics at a site and the interaction of the subpopulations in this site with its environment. The local dynamics at a site were described in the previous section. They are affected by its environment through the inflow of cooperators and defectors. When a site is empty, this inflow determines when it will become inhabited by a cooperator; and when a site is inhabited by cooperators with no defectors, this inflow will determine how long it will take for it to be invaded by a defector—t_f. The environment, on the other hand, is generated by the local dynamics at sites. Essentially, the more intricate are the population dynamics at a site, the more complex is the analysis of the global dynamics. In the d.d. models, the rough temporal structure of the life cycle is rather simple, and it is possible to approximate its global behavior by dividing the life cycle into three main stages: *empty*, *cooperation* and *defection*, where in each the population could be considered as being in one state. In each of these states, the internal dynamics at each site and its interactions with its environment can be described as a Markov process switching between states. Note that this simplified model is very similar to Maynard Smith's (Maynard-Smith, 1974) predator–prey model described in the introduction. For a simplified model of this nature we can obtain an analytical approximation of the conditions for the existence of a non-trivial steady state. We briefly outline a simplified model and its analysis in Sections 4.1, 4.2 and in Appendix C. The life cycle for the d.i. model described in Figs 1 and 5 consists of many different states, since each population structure—n_c, n_d, affects the evolution at the site and the sites effect on its environment differently. Each of these states is characterized by a different outflow of cooperators and defectors. This makes the global analysis of these models much more complicated. For this reason, we restrict the study of their global behavior to simulations. The results from the analysis and simulations of the simplified d.d. model and from the simulations of the d.i. model are presented in Section 4.3.

4.1. A SIMPLIFIED DENSITY-DEPENDENT MODEL

The life cycle at a site for the simplified d.d. model is described in Fig. 6. In this model, for which the life cycle is a simplification of the d.d. life cycle shown in Fig. 2, when a cooperator enters an empty site it immediately establishes a population of n_c cooperators. After some time, the population is invaded by a defector that is either a mutant from within or a migrant from without. Once a defector invades, it instantaneously takes over and establishes a constant population of n_d defectors. This population has a probability P_d per unit time to die and leave the site empty. Diffusion and mutation are stochastic as in the non-simplified models.

The simplified model can be seen as an interacting particle system, where a site (i, j) (which corresponds to the particle) can be in one of the three states: *empty* ($S_{ij} = e$), *cooperation* ($S_{ij} = c$) and *defection* ($S_{ij} = d$). The dynamics of this system can be described as a Markov process, written here in terms of the transition probabilities for a site (i, j) during a time step Δ:

$$A^{ij}_{e\to c}(t) \equiv P(S_{ij}(t+\Delta) = c \,|\, S_{ij}(t) = e) = \frac{D}{4}\, n_c I^{ij}_c(t), \tag{14}$$

$$A^{ij}_{c\to d}(t) \equiv P(S_{ij}(t+\Delta) = d \,|\, S_{ij}(t) = c) = \frac{D}{4}\, n_d I^{ij}_d(t) + \mu n_c, \tag{15}$$

$$A^{ij}_{d\to e}(t) \equiv P(S_{ij}(t+\Delta) = e \,|\, S_{ij}(t) = d) = P_d \tag{16}$$

$$A^{ij}_{e\to e}(t) = 1 - A^{ij}_{e\to c}(t), \tag{17}$$

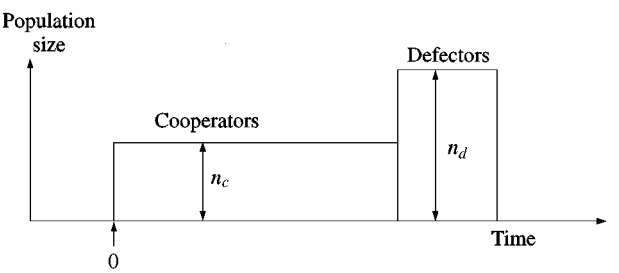

FIG. 6. A life cycle in the simplified model: population as a function time. Notice the similarity to the density-dependent life cycle (Fig. 2).

$$A^{ij}_{c\to c}(t) = 1 - A^{ij}_{c\to d}(t),$$

$$A^{ij}_{d\to d}(t) = 1 - A^{ij}_{d\to e}(t). \qquad (18)$$

Here, the number of (i, j)'s nearest neighbors in state c is denoted as I^{ij}_c, and the number of nearest neighbors in state d is denoted as I^{ij}_d. In writing these dynamics it was assumed that the time step $\Delta \ll 1$, so that effects that are second order in μ and D can be ignored. Equation (14) describes how a site changes its state from e to c, by way of a nearest-neighbor interaction corresponding to diffusion. Equation (15) describes how a site changes its state from c to d, either by nearest-neighbor interaction corresponding to diffusion, or spontaneously in a way which corresponds to mutation. Finally, eqn (16) describes how a site changes its state from d to e, spontaneously, in a way that corresponds to the death of the defector population.

Equations (14)–(18) have four parameters: Dn_c, Dn_d, μn_c and P_d. As D, μ, and P_d are all homogeneous to the first order in the time-scale Δ, so are the right-hand sides (r.h.s.) in eqns (14)–(18). This means that one of the four parameters, such as P_d, could be taken to determine the time-scale. The other three could be taken to be independent of the time-scale, for example, Dn_d/P_d, D/μ and n_c/n_d (which are independent and homogeneous with degree 0 in Δ).

This model captures the qualitative features of the local behavior of the explicit d.d. models of Section 2. The establishment of cooperation, the defector take-over and the populations' extinction which derive from the population dynamics at a site in the explicit d.d. models are assumed in the simplified model. The interactions between a site and its environment, however, are of the same form in both simplified and general d.d. models. The environment affects when the empty site becomes inhabited and when the defector take-over occurs. On the other hand, a site affects its environment by diffusing out cooperators and defectors.

4.2. A MEAN-FIELD APPROXIMATION TO THE SIMPLIFIED DENSITY-DEPENDENT MODEL

We would like to find the region in the model's parameter space in which a non-trivial steady state exists. One way to do this, is to solve the model analytically. A solution is a stationary probability distribution on the space of all possible lattice configurations $P(\{S_{ij}\}_{i,j\in Z})$ as a function of the model's parameters. Using a mean-field approximation one can find the best solution within a restricted class of distributions. Roughly speaking, as the class of distributions becomes larger the approximations become better. In this paper, we will not evaluate the accuracy of the approximations other than by comparing their predictions with simulations. A systematic evaluation of these approximations, as well as a more accurate analysis using Renormalization Groups, has been done for other particle systems (Goldenfeld, 1992; Baxter, 1982).

The first-order mean-field approximation is restricted to probability distributions of the form

$$P(\{S_{i,j}(t)\}_{i,j\in Z}) = \prod_{i,j\in Z} P(S_{i,j}(t)). \qquad (19)$$

This means that the probability of finding the system in a certain configuration can be decomposed into a product of the probabilities of finding each site in its state. One further assumes that the probabilities of finding a site in state c, d or e are uniform across the lattice. Under these assumptions the system's description reduces to the probabilities of finding any site in each one of the possible states.‡ Denoting these probabilities which are independent of the site by p_e, p_c, and p_d, the system's dynamics reduces to

$$p_e(t+\Delta) = p_e(t)(1 - A_{e\to c}(t)) + p_d(t)A_{d\to e}(t),$$

$$p_c(t+\Delta) = p_c(t)(1 - A_{c\to d}(t)) + p_e(t)A_{e\to c}(t),$$

$$p_d(t+\Delta) = p_d(t)(1 - A_{d\to e}(t)) + p_c(t)A_{c\to d}(t), \qquad (20)$$

where $A_{e\to c}$, $A_{c\to d}$ and $A_{d\to e}$ denote the transition probabilities, which can be derived from eqns (14):

$$A_{e\to c}(t) = \frac{D}{4} n_c 4p_c(t),$$

‡ Note, that both this and the "second-order" mean-field approximations, can be treated as models for the persistence of cooperation by their own right.

$$A_{c\to d}(t) = \frac{D}{4}\, n_d 4p_d(t) + \mu n_c,$$

$$A_{d\to e}(t) = P_d. \tag{21}$$

Here we set $I_d = 4p_d(t)$ and $I_c = 4p_c(t)$.

The fixed point and stability analysis for this system is straightforward. A non-trivial fixed point (one where $p_e \neq 1$) exists if

$$\frac{D}{\mu} > 1. \tag{22}$$

When this condition holds, the system has two meaningful fixed points, with one being trivial ($p_e = 1$) and the other not. In this case, only the non-trivial fixed point is stable, and thus the persistence of cooperation is obtained for any initial condition in which $p_c(0) \neq 0$. For this non-trivial fixed point, expressions for R, M or any other dynamic parameter of the system, as functions of Dn_d/P_d, D/μ, n_c/n_d and P_d can be derived.

Deriving condition 22 from general considerations will help in understanding the scope of the first-order approximation. For a non-trivial steady state to be maintained, every life cycle has on average to establish exactly one new life cycle. This requirement takes the form:

$$DS_c\rho_e^c = \frac{M}{1 + 1/R}\,\rho_e^c = 1, \tag{23}$$

where ρ_e^c denotes the density of empty sites near a site in state c. This density equals the probability that a cooperator leaving a site will establish a new life cycle. Condition (22) could be derived from eqn (23) by putting trivial bounds on ρ_e^c and S_c:

$$\rho_e^c \leqslant 1, \tag{24}$$

$$S_c = n_c t_f \leqslant n_c \frac{1}{\mu}. \tag{25}$$

The bound on ρ_e^c is realized only when all the neighboring sites are empty. The bound on t_f is also realized when all the neighboring sites are empty, i.e. when the first defector is always a mutant. These two bounds imply that condition (22) is equivalent to the requirement that at least one cooperator diffuses out in a life cycle at a site surrounded by empty neighbors. As the number of cooperators in a life cycle at an isolated site depends only on μ, the number of cooperators diffusing from it depends only on μ and D.

Condition (22) indicates that the first-order mean-field approximation cannot incorporate the harmful effects of migration, an important feature of the model. In the first-order mean-field approximation, the density ρ_e^c can approach 1 enabling cooperators to survive as long as on average one cooperator migrates during a life cycle. This means that in this approximation defector migration does not really determine whether cooperation prevails or not, because the density of inhabited sites can always be so low that no defector ever invades it. In the spatial model the density ρ_e^c can never reach 1, because near a population of cooperators there is always a finite probability of having the population from which the founding cooperator migrated. The neighboring population, in this case, will be in either state c or d during some part of the life cycle of its daughter subpopulation. This discrepancy between the spatial model and the first-order mean-field approximation, is demonstrated in Figs 3 and 4. The second picture in Fig. 4 (25 generations) indicates that a life cycle at a site surrounded by empty sites produced more than one diffusing cooperator. Yet cooperation does not prevail due to the effects of extensive defector migration into sites inhabited by cooperators. An approximation incorporating such effects would have to describe the correlations between the states of nearest neighbors. Such an approximation is outlined in Appendix C. Results presented in the next section will hint at the possibility that as the phase transition between persistence and non-persistence of cooperation is approached the correlation length in the system goes to infinity. This would imply that near the parameters at which the transition happens, the reliability of such mean-field approximations is questionable.

4.3. RESULTS I: REGIONS CHARACTERIZED BY THE DYNAMICAL PERSISTENCE OF COOPERATION

The d.i. model has six parameters: α, β, γ, δ, μ and D, while the simplified d.d. model has four: D/μ, Dn_d/P_d, n_c/n_d and P_d. A point at which

a non-trivial steady state is maintained, is characterized by a stationary probability distribution on all possible lattice configurations. We would like to present these (6/4)-dimensional phase spaces in a comprehensible way, that permits comparison with the global behavior of models which derive from different local parameters. In doing so, we will necessarily lose some information, information that can be further explored using different representations. To the extent that the fine temporal and spatial structure of a steady state in these models can be ignored, the basic variables characterizing the global dynamics would be M—the average number of migrants during a life cycle, and R—the cooperator to defector ratio among these migrants.

Phase spaces in the R–M coordinates, which were derived from analytical approximations to the simplified d.d. model, from simulations of the simplified d.d. model and from simulations of the d.i. model are presented in Figs 7 and 8. The solid lines in Figs 7 and 8 correspond to the boundaries in R–M space below which a non-trivial steady state does not exist according to the first and second-order approximations. We will refer to such a boundary as a *phase boundary*. The thick lines in Figs 7 and 8 represent the phase boundaries derived from simulations. They were derived as described in Appendix D. In the simplified d.d. model, $DS_d = Dn_d/P_d$ is one of the basic parameters of the model, while DS_c derives from the dynamics (see Sections 3.2 and 4.1). As $M = DS_d + DS_c$ and $R = DS_c/DS_d$, one component of R and M is a parameter whereas the other is an outcome of the dynamics depending on the other parameters. In the d.i. model the situation is similar (3.1), R is a function of α, β, γ and δ, and thus can be considered to be a parameter, whereas M derives from the dynamics which depends on the other parameters.

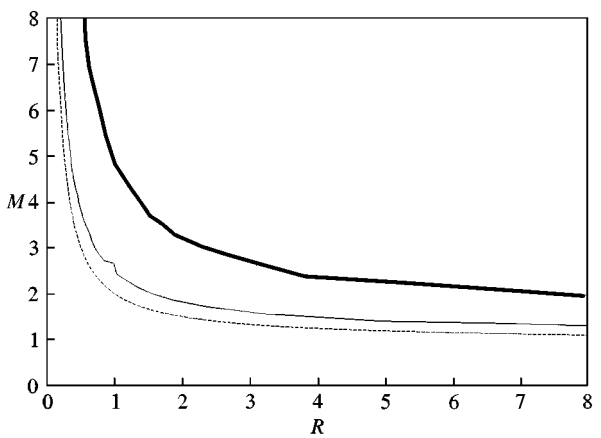

FIG. 7. The R–M phase space for the simplified d.d. model: The graph presents regions of cooperation persistence according to the first and second-order mean-field approximations (above the phase boundaries pictured); and according to simulations. For details on how the phase boundaries were derived from simulations, see Appendix D: (——) second order; (·····) first order; (▬▬) simplified model.

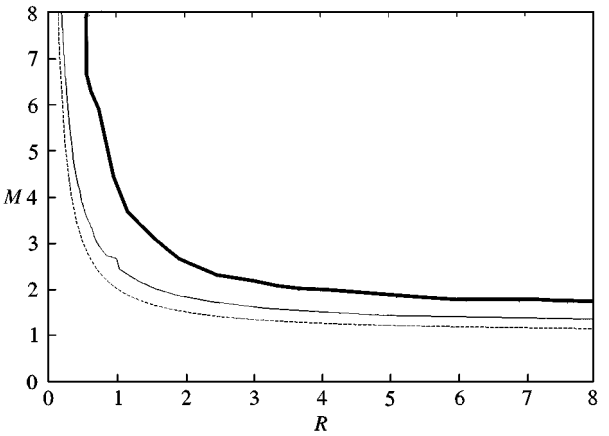

FIG. 8. The R–M phase space for the d.i. model: The graph presents regions of cooperation persistence according to simulations of the d.i. model. The phase boundaries according to the mean-field approximations to the simplified d.d. model are also presented, as reference. For details on how the phase boundaries were derived from simulations, see Appendix D: (——) second order; (·····) first order; (▬▬) d.i. model.

The shape of the phase boundaries from the analysis and simulations can be roughly understood from the heuristic derivation in the previous section, eqn (23), which predicts that the phase boundary takes the form:

$$M = C\left(1 + \frac{1}{R}\right), \tag{26}$$

where C is some constant. The differences in the shape and position of the phase boundaries reflect the effects of the fine spatio-temporal dynamic structure. As we discussed at the end of the previous section, one can state roughly that the effect of spatial correlations, i.e. spatio-temporal structure, is to increase the damage that

defectors inflict—thus imposing stronger restrictions on the region in the R–M space where a non-trivial steady state can be maintained. This causes the phase boundaries resulting from the simulations to be above those resulting from the second-order approximation; as well as for second-order phase boundary to be above the first order. A systematic study of the factors effecting the phase boundaries requires the study of higher-order correlations.

4.4. RESULTS II: HOW LOWER INDIVIDUAL FITNESS INDUCES HIGHER SURVIVABILITY

The drawback of using R–M phase spaces to study a specific model is a loss of information about the relation between the system's behavior and its basic parameters. In the d.i. model the persistence of cooperation depends on the parameter α—the fitness associated with interactions between cooperators. In Section 3.1 we explain why a reduction in α leads to a larger cooperator to defector ratio R, and derive the functional dependence of R on α. As α increases and R decreases, we expect that the disturbance from defectors will grow to a point where cooperation cannot be maintained. This effect is illustrated in Figs 3 and 4. Across two systems we took all the parameters other than α to be equal. The system with the smaller α reached a non-trivial steady state in which cooperation persisted, whereas the system with the larger α reached the trivial steady state without cooperation.

Figs 9(a)–(c) illustrate the behavior of several dynamic variables as a function of α, while all the other parameters are fixed. When α increases, the number of migrating defectors also increases while the number of migrating cooperators remains approximately constant [Fig. 9(a)]. Thus, both R [Fig. 9(b)], and the density of occupied sites [Fig. 9(c)] decrease. Hence, a decrease in the individual fitness of cooperators leads to increases in: the total number of cooperators in the system, the density of sites they occupy, and their numbers relative to defectors. Note that the measured R [Fig. 9(b)] is very close to the analytically derived value. This supports the scaling argument described in Section 3.1 and demonstrated in Fig. 5.

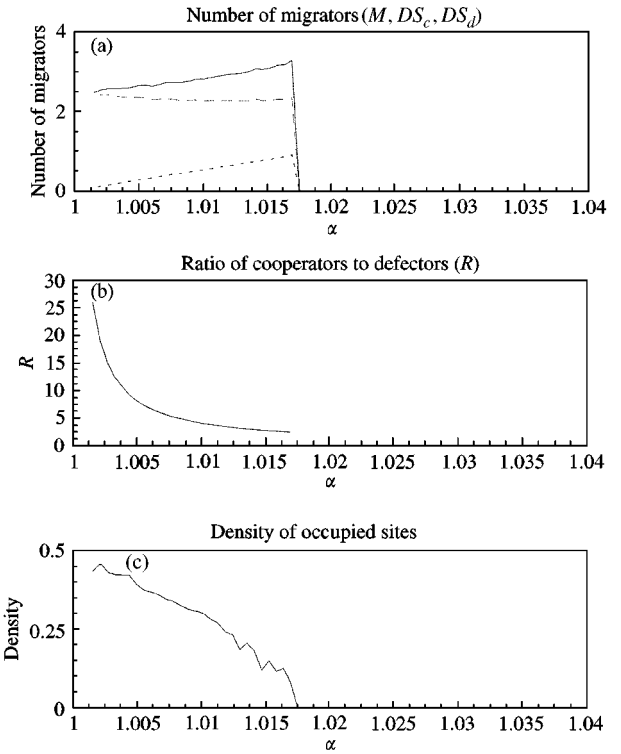

FIG. 9. Dynamic behavior in the d.i. model as a function of α: Figs (a)–(c) describe different dynamic variables of the system as they result from simulations of the d.i. model with $\beta = 0.95$, $\gamma = 0.8$, $\delta = 1.2$, $\mu = 0.0001$ and $D = 0.0003$ where α varies between 1.001 and 1.035. (a) The average number of migrating cooperators (DS_c), defectors (DS_d) and their sum (M) in a life cycle given as a function of α. (b) the average ratio of cooperators to defectors migrating during a life cycle (R) is given as a function of α; and (c) the density of occupied sites is given as a function of α. (·····) M; (– – –) DS_c; (- - -) DS_d.

Around $\alpha = 1.0175$, the rate of destruction by migrating defectors reaches a level that precludes the maintenance of a non-trivial steady state, like the example described in Fig. 4. Increasing the individual fitness of cooperators therefore leads to a condition in which cooperation can no longer persist. Note that near the phase boundary the density of occupied sites drops [Fig. 9(c)]. This suggests that the correlation distance in the system grows at this vicinity. As mentioned in the previous section, this casts doubts on the reliability of the mean-field approximations near the phase boundary.

5. Discussion

We have demonstrated that cooperation may persist in a dynamic mode where populations of

cooperators and defectors constantly appear and disappear. We explain the persistence of cooperation by considering the life cycle of a population at a site. The life cycle starts when one cooperator establishes a population; this population grows; defectors invade and take over; and ultimately the population goes extinct. During this life cycle, new populations of cooperators are founded by single cooperators that migrate to empty neighboring sites. The system reaches a steady state where cooperation persists, if the global "birth" rate of populations is equal to their "death" rate, or equivalently, if on average every population gives rise to one other population during its life cycle. This steady state arises from a repeated turnover of populations—cooperation persists although every single population of cooperators eventually dies out. In Section 4, we demonstrate that these dynamics enable the persistence of cooperation in a large section of the model's parameter space. Furthermore, we demonstrate and explain that lowering the local fitness of cooperators in the d.i. model can enable the persistence of cooperation. Within the region of persistence, lowering the local fitness of cooperators can increase the number of cooperators, the density of sites inhabited by them, and their numbers relative to the defectors.

The dynamic mode we have described may appear in a variety of biological systems. In addition to Maynard-Smith's predator–prey model reviewed in the introduction, we consider one model for the persistence of altruism and one model for the persistence of "prudent" predation in a predator–prey system. Epstein (1998) observes oscillatory behavior in a spatial PD model which is analogous to the dynamic mode that we describe. In this model, individuals which are either cooperators or defectors occupy sites on a two-dimensional lattice. Each individual plays the PD game with its neighbors, where the payoffs it accumulates determine its probability to produce an identical offspring or to die and leave its site empty. In one version, he sets the payoffs for cooperator–cooperator interactions to be positive and the payoffs for defector–defector interaction to be negative. This results in oscillation of the total number of cooperators and defectors over time, where the peaks in defectors appear to closely follow the peaks in cooperators. Epstein also notes that in this dynamical regime, decreasing the payoffs for cooperator–cooperator interactions may improve the cooperators to defectors ratio. This is similar to the mode that we describe, where the localized subpopulations in our model are analogous to the spatially extended neighborhoods in Epstein's model.

Gilpin (1975) studies a model for the persistence of altruistic behavior in the context of predator–prey systems. Through computer simulations of structured populations, in which subpopulations of predators and prey inhabit isolated patches, he studies the persistence of predator "prudence" restraint to not over-exploit their food supply. In his model, "selfish" predators in a patch drive the prey to extinction, which in turn drives the predator population in the patch to extinction. Gilpin allows for migration between patches and genetic drift within them. Although he does not find parameter values where the "selfish" and "prudent" predators can coexist, we believe that introducing mutations that cause the "selfish" predators to reappear, or increasing the number and perhaps introducing spatial organization of patches would result in the mode of dynamic coexistence we have described.

Maynard-Smith (1976) considers the implications of migration in a patchy environment on the persistence of altruism. He suggests a criterion based on the average number of new defector populations founded by migration from a patch with defectors before that patch goes extinct, which he denotes as M.§ In the following summary of his reasoning, note that in our system, M should be defined in terms of cooperator populations rather than defector populations.‖

§ Note that we have used M for a different meaning.

‖ Maynard-Smith considers systems where defectors can either persist in coexistence with cooperators or go extinct. In our system these options apply to cooperators rather than to defectors. Defectors and cooperators are not equivalent in our system because we have incorporated mutation and defectors may depend on repeated appearance via mutation in order to persist. We note, however, that even in the absence of mutation there are parameter regions in our system where cooperators and defectors coexist in a steady state. In these cases, M can be defined for either cooperators or defectors.

He claims that if $M > 1$, then defectors would take-over the population, while if $M < 1$, then cooperators would prevail and defectors would go extinct. If we define M for cooperators rather than defectors, then $M = 1$ corresponds to our steady state. From Maynard-Smith's formulation, one may assume that the case $M = 1$ is a mathematical artifact requiring exact parameters and therefore it should not be considered seriously. This is not true for our system, however, and it need not be true for other systems. In our system, the value of this variable M, which is defined for cooperators rather than defectors, is produced by the dynamics of the system. Within the parameter range where cooperation persists, changes in the underlying parameters affect other variables of the system, such as the density of sites that are empty or inhabited by cooperators or the average number of cooperators in the system (see Fig. 9), but leave $M = 1$. One mechanism underlying this stability, holds when the parameters of the system are changed such that the extinction of subpopulations becomes faster and the number of migrating cooperators becomes smaller. Under these conditions the density of empty sites may increase thus increasing the chance of a migrating cooperator to colonize an empty site (see Fig. 9). A similar process stabilizes the steady state with fixed parameters: when the density of sites occupied by cooperators drops below the steady state level, a migrating cooperator has an increased probability of finding an empty site. A more precise structural and dynamical stability analysis would consider spatio-temporal patterns such as the correlation between newly inhabited sites with neighboring sites inhabited by defectors. We note, however, that the mode of persistence that we describe seems both structurally and dynamically stable, and that this may result from self-regulating processes of the type discussed above. Maynard-Smith reaches a similar conclusion in the analysis of the predator–prey model described in the introduction to this paper. The extent to which the dynamic mode we have described occurs in biological systems depends on the parameter values in these systems. Yet the fact that this mode is both structurally and dynamically stable and holds for a large range of parameters, makes it likely to occur in natural systems.

We describe how cooperation persists, but not how it originates in the first place, or continues to evolve once it is established. Although a proper treatment of these questions requires extensions to our model, we offer a few comments here. When considering the origin and evolution of cooperative behavior, one should remember that cooperative and defective behaviors are often relative terms. In a homogeneous population, the appearance of an individual behaving more altruistically than its peers may elicit a dynamic where the pre-existing type is redefined as a defector. If cooperation is costly, the pre-existing type benefits from interacting with the new type without having to pay the cost, and will have the characteristics of defectors upon the appearance of cooperators.

We suggest two scenarios for the origination of the mode we have described. Consider a system consisting of reproductively isolated subpopulations, where new subpopulations are founded by individuals that leave existing subpopulations. Assume further that individuals with cooperative behavior appear in the system, after which defectors, which may have been the pre-existing type, appear. The type of dynamic we describe requires one additional condition: subpopulations of defectors must not be self-maintaining. This condition seems unreasonable at first, if defectors were the pre-existing type and therefore existed independently of cooperators. We suggest two ways to account for this seeming paradox. One is to consider a system with an inhomogeneous environment containing harsh areas where subpopulations of the pre-existing type are not able to survive. Since cooperator subpopulations are more efficient, they can inhabit some of these niches. Once they do, the pre-existing type may invade these areas, by taking advantage of the cooperators, and thus the conditions for the dynamic persistence of cooperation may arise upon invasion by cooperators. Another possibility is a system in which reproductively isolated subpopulations share common resources. Once cooperating subpopulations appear, the conditions for a solitary subpopulation change for the worse, again due to the higher efficiency of subpopulations consisting of cooperators. Consequently, subpopulations of defectors are not self-maintaining,

and the stage is set for a steady state of the type we describe.

Now consider the course of evolution after a population reaches a steady state with cooperation. Again we assume that subpopulations share common resources. At some point "improved" cooperators appear, and consequently both the old cooperators and defectors assume the role of defectors. The long-term evolution (Eshel *et al*., 1997, 1998) of the system then becomes relevant. Specifically, one should consider the conditions for invasion by the new cooperators. On the one hand, "improved" cooperators may invade by establishing the conditions, through the shared resources, under which subpopulations of the preceding cooperators cannot maintain themselves. If this is the case, then after the new cooperators take-over, the population reaches a new steady state in which the carrying capacity of the environment has increased, and the fitnesses within subpopulations are renormalized.

Taking a higher cooperator fitness while leaving everything else the same can lead, however, to a breakdown in the persistence of cooperation, as we describe in Section 4. If an improved cooperator is characterized by a large α when it appears in a given patch, then a large number of defectors is generated during a life cycle, which annihilates the population in its vicinity (see Fig. 9). This leads to the extinction of the "improved" cooperators and to the continuing persistence of the pre-existing cooperators. Therefore, invasibility conditions for "improved" cooperators in these systems can be subtle and deserve a closer analysis. We conjecture that such factors dictate the rate at which cooperation evolves, and may prevent it from improving significantly in a single transition. If this is true, it will be reflected in both the invasion criterion, and in the renormalization of fitness after a take-over event.

This work originated while we were staying at the home of Arian and Amnon Tamir in Paris, in the summer of 1995. We thank them for their kind hospitality. We have benefited greatly from comments of, and discussions with, Ilan Eshel. We are also grateful for editorial corrections suggested by Lauren Ancel, Susan Ptak, and Carl Bergstrom. Last but not least we thank Marcus W. Feldman and the entire Feldman lab for their help. This work was supported by NIH grant GM 28016 to M. W. Feldman, and by the Santa Fe Institute.

REFERENCES

AXELROD, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

BAXTER, R. J. (1982). *Exactly Solved Models in Statistical Mechanics*. London, New York: Academic Press.

COHEN, D. & ESHEL, I. (1976). Founder effect and evolution of altruistic traits. *Theoret. Popul. Biol.* **10,** 276–302.

COOPER, B. & WALLACE, C. (1998). Evolution, partnership and cooperation. *J. theor. Biol.* **195,** 315–328.

EPSTEIN, J. (1998). Zones of cooperation in demographic prisoner's dilemma. *Complexity* **4,** 36–48.

ESHEL, I. (1971). On the neighbor effect and the evolution of altruistic traits. *Theoret. Popul. Biol.* **3,** 258–277.

ESHEL, I. & CAVALLI-SFORZA, L. L. (1982). Assortment of encounters and evolution of cooperativeness. *Proc. Natl. Acad. U.S.A.*, **79,** 1331–1335.

ESHEL, I., FELDMAN, M. W. & BERGMAN, A. (1998). Long-term evolution, short-term evolution, and population genetic theory. *J. theor. Biol.* **191,** 391–396.

ESHEL, I., MOTRO, U. & SANSONE, E. (1997). Continuous stability and evolutionary convergence. *J. theor. Biol.* **185,** 333–343.

FERRIERE, R. & MICHOD, R. E. (1995). Invading wave of cooperation in a spatial iterated prisoner's dilemma. *Proc. Royal Soc. Lond. B* **259,** 77–83.

GILPIN, M. E. (1975). *Group Selection in Predator–prey Communities*. Princeton, NJ: Princeton University Press.

GOLDENFELD, D. (1992). *Lectures on Phase Transitions and the Renormalization Group*. Reading, MA, Addison-Wesley.

HAMILTON, W. D. (1964). The genetical evolution of social behavior, i and ii. *J. theor. Biol.* **7,** 1–52.

HUANG, K. (1987). *Statistical Mechanics*. New York: John Wiley & Sons.

MAYNARD-SMITH, J. (1964). Group selection and kin selection. *Nature* **201,** 1145–1147.

MAYNARD-SMITH, J. (1974). *Models in Ecology*. U.K.: Cambridge University Press.

MAYNARD-SMITH, J. (1976). Group selection. *Quart. Rev. Biol.* **51,** 277–283.

NAKAMARU, M. H. M. & IWASA, Y. (1997). The evolution of cooperation in a lattice-structured population. *J. theor. Biol.* **184,** 65–81.

NAKAMARU, M. H. M. & IWASA, Y. (1998). Score-dependent fertility model for the evolution of cooperation in a lattice. *J. theor. Biol.* **194,** 101–124.

NOWAK, M. A. & MAY, R. M. (1992). Evolutionary games and spatial chaos. *Nature* **359,** 826–829.

OLIPHANT, M. (1999). Evolving cooperation in the non-iterative prisoner's dilemma: the importance of spatial organization. In: *Artificial Life IV* (Brooks, R. & Maes, P., eds). Cambridge, U.S.A.: MIT University Press.

ROBERTS, G. (1998). Competitive altruism: from reciprocity to the handicap principle. *Proc. Roy. Soc. Lon. B* **265,** 427–431.

WEIBULL, W. J. (1995). *Evolutionary Game Theory*. Cambridge, MA: MIT University Press.

WILSON, D. S. (1987). Altruism in mendelian populations derived from sibling groups: the hey stack model revisited. *Evolution* **41,** 1059–1970.

APPENDIX A

Proof of Result 1

Result 1. *In the density-independent model the ratio R of the average number of migrating cooperators to the average number of migrating defectors in a life cycle is*

$$R(\alpha, \beta, \gamma, \delta, \mu, D) = \frac{E(M_c)}{E(M_d)} = \frac{DE(S_c)}{DE(S_d)} = \frac{1-\beta}{\alpha-1}\frac{\alpha-\gamma}{\delta-\beta} + O[\mu, D]. \tag{A.1}$$

Proof. The life cycle begins when the first cooperator enters an empty size. Denoting this time as $t = 0$, we have

$$n_c(0) = 1, \tag{A.2}$$

$$n_d(0) = 0. \tag{A.3}$$

The population at a site then begins evolving according to eqns (4). These equations can be written as follows, separating zero- and first-order terms in μ and D:

$$E(n_c(t+1)|n_c(t), n_d(t)) = f_c(n_c(t), n_d(t))n_c(t) + \left[\mu\left(f_d(n_c(t), n_d(t))n_d(t) - f_c(n_c(t), n_d(t))n_c(t)\right) + D\left(\frac{1}{4}\sum_{n.n.} n_c^{n.n.}(t) - f_c(n_c(t), n_d(t))n_c(t)\right)\right] \tag{A.4}$$

$$E(n_d(t+1)|n_c(t), n_d(t)) = f_d(n_c(t), n_d(t))n_d(t) + \left[\mu\left(f_c(n_c(t), n_d(t))n_c(t) - f_d(n_c(t), n_d(t))n_d(t)\right) + D\left(\frac{1}{4}\sum_{n.n.} n_d^{n.n.}(t) - f_d(n_c(t), n_d(t))n_d(t)\right)\right]. \tag{A.5}$$

The first-order terms in μ and D affect the dynamics in two ways: the first is by slightly changing the population sizes due to migration between sites and mutation between types. This changes R to the first order in μ and D. The second is by affecting t_f, the time when the first defector appears in the life cycle. The appearance of the first defector has a dramatic effect on the life cycle, as it marks the beginning of defector takeover. Consequently, we will ignore the effects of mutation and migration at all times other than when the defector population size is 0. This will be done by incorporating a "source" to the zero-order defector dynamics, which is "on" as long as $n_d = 0$ and "off" otherwise. This source term adds one defector at time $t+1$ with the same probability with which it would appear as a result of cooperator mutation and defector migration.

The zero order (in μ and D) dynamics with the source term are then described by

$$E(n_c(t+1)|n_c(t), n_d(t)) = f_c(n_c(t), n_d(t))n_c(t), = \left(\alpha\frac{n_c(t)}{n(t)} + \gamma\frac{n_d(t)}{n(t)}\right)n_c(t) \tag{A.6}$$

$$E(n_d(t+1)|n_c(t), n_d(t)) = f_d(n_c(t), n_d(t))n_d(t) + I(n_d(t))P(n_d(t+1) \neq 0|n_c(t), n_d(t)) = \left(\delta\frac{n_c(t)}{n(t)} + \frac{n_d(t)}{n(t)}\right)n_d(t) + I(n_d(t))P(n_d(t+1) \neq 0|n_c(t), n_d(t)), \tag{A.7}$$

where

$$I(n_d(t)) = \begin{cases} 1 & n_d(t) = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{A.8}$$

Denoting $a \equiv (\alpha - \gamma)$ and $b \equiv (\delta - \beta)$ these equations could be written as

$$E(n_c(t+1)|n_c(t), n_d(t)) = \alpha n_c(t) - a\frac{n_c(t)n_d(t)}{n(t)}, \tag{A.9}$$

$$E(n_d(t+1)|n_c(t), n_d(t)) = \beta n_d(t) + b\frac{n_c(t)n_d(t)}{n(t)} + I(n_d(t))P(n_d(t+1) \neq 0|n_c(t), n_d(t)). \quad \text{(A.10)}$$

Taking a linear combination of eqns (A.9) to eliminate the nonlinear term gives

$$bE(n_c(t+1)|n_c(t), n_d(t)) + aE(n_d(t+1)|n_c(t), n_d(t)) = b\alpha n_c(t) + a\beta n_d(t) + aI(n_d(t))P(n_d(t+1) \neq 0|n_c(t), n_d(t)). \quad \text{(A.11)}$$

In order to turn the conditional averages and free variables to averages, we multiply eqn (A.11) by $P(n_c(t), n_d(t))$ and sum over all possible values of $n_c(t)$ and $n_d(t)$, to get

$$bE(n_c(t+1)) + aE(n_d(t+1)) = b\alpha E(n_c(t)) + \alpha\beta E(n_d(t)) + aP(n_d(t+1) \neq 0; n_d(t) = 0), \quad \text{(A.12)}$$

where we used

$$\sum_{n_c(t), n_d(t)} I(n_d(t))P(n_d(t+1) \neq 0|n_c(t), n_d(t))P(n_c(t), n_d(t)) = \sum_{n_c(t)} P(n_d(t+1) \neq 0|n_c(t), n_d(t) = 0) = P(n_d(t+1) \neq 0, n_d(t) = 0). \quad \text{(A.13)}$$

Summing these equations from $t = 0$ to ∞, we get

$$b(E(S_c) - 1) + aE(S_d) = b\alpha E(S_c) + a\beta E(S_d) + a, \quad \text{(A.14)}$$

because

$$\sum_{t=0}^{\infty} P(n_d(t+1) \neq 0; n_d(t) = 0) = 1. \quad \text{(A.15)}$$

This sum is simply the probability that the first defector will appear sometime (ignoring cases in which defectors disappear and then appear again). Reorganizing eqn (A.14) and re-substituting a and b, we get

$$R = \frac{E(S_c)}{E(S_d)} = \frac{1-\beta}{\alpha-1}\frac{\alpha-\gamma}{\delta-\beta} - \frac{1}{E(S_d)}\frac{\alpha+\delta-\beta-\gamma}{(\alpha-1)(\delta-\beta)}. \quad \text{(A.16)}$$

It remains to be shown that $1/E(S_d) \approx O[\mu, D]$. We will not give a complete formal proof, instead we provide the essence of the argument. We begin by showing that $1/E(S_c) \approx O[\mu, D]$. Assuming that the first defector appears as a result of a mutation, it would appear when $\int_0^{t_f} n_c(t)\mathrm{d}t \approx 1/\mu$, therefore $S_c \geqslant 1/\mu$. If on the other hand, it appears as a result of diffusion then $t_f \approx 1/D\langle n_d\rangle$ where $\langle n_d\rangle$ is the average number of defectors in a site's neighborhood. In this case, $\int_0^{t_f} n_c(t)\,\mathrm{d}t \approx \alpha^{t_f} \approx \alpha^{1/D\langle n_d\rangle}$; therefore, $S_d \geqslant O[1/D]$. If the probability of the first defector appearing from the two processes is comparable, both estimates are valid. Hence, we conclude that $E(S_c) \approx O[1/\mu, 1/D]$, and since S_d scales with S_c we get $1/E(S_d) \approx O[\mu, D]$ which concludes the proof that

$$R(\alpha, \beta, \gamma, \delta, \mu, D) = \frac{DE(S_c)}{DE(S_d)} = \frac{1-\beta}{\alpha-1}\frac{\alpha-\gamma}{\delta-\beta} + O[\mu, D]. \quad \text{(A.17)}$$

APPENDIX B

Proof of Result 2

In Section 3.2, we considered the life cycle in a specific d.d. model. We found that for any α, β, γ, and δ, R could be made larger than any R^* provided that μ and D are taken to be small enough. An expression relating R^* to μ and D was given. In this appendix, that expression will be derived for a class of d.d. referred to as d.d. models of type 1:

Definition. A model with absolute fitness functions of the form

$$f_c(n_c, n_d) = g(n)\left(\alpha\frac{n_c}{n} + \gamma\frac{n_d}{n}\right),$$

$$f_d(n_c, n_d) = g(n)\left(\delta\frac{n_c}{n} + \beta\frac{n_d}{n}\right),$$

will be called a *density-dependent model of type* 1, if

1. $g(n) \leqslant 1$ for all n.
2. The average population size at each time step is bounded by some maximum population size n_{max}. This means that for all $n_c(t)$ and $n_d(t)$:

$$E(n(t+1)|n_c(t), n_d(t)) \leqslant n_{max}. \quad \text{(B.1)}$$

3. For $\alpha > 1$ there exists a time $t(\alpha)$ and a constant $0 < c(\alpha) \leqslant 1$ such that $E(n_c(t)) \geqslant c(\alpha)n_{max}$ for $t > t(\alpha)$ when there are no defectors, taking $n_c(0) = 1$.

The first requirement means that the population size in each time step is bounded from above by the population resulting from the density-independent dynamics. The second condition is a requirement for limiting the population size. It is surely satisfied by requiring $g(n) \leqslant \min\{n_{max}/\delta n, 1\}$ for all n, since then we have

$$E(n(t+1)|n_c(t), n_d(t)) \leqslant g(n)\,\delta n(t) \leqslant n_{max}.$$

Both conditions are trivially met by the function $g(n)$ described at the beginning of Section 3.2. The third requirement captures the way that the cooperator population assumes a certain size after time $t(\alpha)$, and remains in this size until t_f. In the example, this condition is satisfied by choosing $c(\alpha) = \alpha/\delta$ and $t(\alpha) = [\log_\alpha n_{max}/\delta] + 1$.

Under these conditions we prove:

Result 2. *Given a density-dependent model of type* 1 *and* $R^* \geqslant 0$, *taking* μ *and* D *such that*

$$D + \mu \leqslant \frac{c(\alpha)}{c(\alpha)n_{max}t(\alpha) + \widetilde{S_d}\,R^*},$$

ensures that R *is bounded from below by* R^*, *where* $\widetilde{S_d}$ *is* $E(S_d)$ *for the density-independent model with the same parameters* α, β, γ, δ, *and initial conditions* $\widetilde{n}_c(0) = n_{max}$, $\widetilde{n}_d(0) = 1$.

Proof. The proof follows the same line of reasoning applied to the example in Section 3.2. By increasing t_f one can increase S_c as much as required, leaving S_d bound. However, increasing t_f simply means taking smaller μ and D.

In order to find the condition that t_f must satisfy, we begin by finding a lower bound for R. We do so by finding a lower bound for $E(S_c)$ and an upper bound for $E(S_d)$:

$$E(S_c) = E\left(\sum_{t=0}^{\infty} n_c(t)\right) \geqslant E\left(\sum_{t=t(\alpha)}^{t_f} n_c(t)\right) \quad \text{(B.2)}$$

$$= \sum_{t=t(\alpha)}^{t_f} E(n_c(t)) \geqslant (t_f - t(\alpha))c(\alpha)n_{max},$$

$$E(S_d) = E\left(\sum_{t_f}^{\infty} n_d(t)\right)$$

$$\leqslant E\left(\sum_{t_f}^{\infty} \widetilde{n}_d(t)|\widetilde{n}_c(0) = n_{max}, \widetilde{n}_d(0) = 1\right) \equiv \widetilde{S_d}, \quad \text{(B.3)}$$

where $\widetilde{n}_d(t)$ is the defector population at time t for the density-independent model with the same α, β, γ and δ, and initial conditions: $\widetilde{n}_c(0) = n_{max}$ and $\widetilde{n}_d(0) = 1$. In bounding $E(S_d)$, we have used condition 1 (from the definition of type 1), which implies that the d.i. population bounds the d.d. population with the same parameters. From eqn (B.3) we get

$$R \geqslant \frac{c(\alpha)n_max}{\widetilde{S_d}}(t_f - t(\alpha)). \quad \text{(B.4)}$$

Next, we find a lower bound for t_f depending on μ and D. The first defector in the life cycle can appear either as a mutant or as a migrant from a neighboring site. The probability for a defector to appear as a mutant in one time step is bounded by μn_{max}, while the probability that a defector would migrate from a neighboring site, could be bounded by $4(\frac{1}{4}Dn_{max}) = Dn_{max}$. From these

bounds we get a lower bound on t_f in terms of μ and D:

$$t_f \geqslant \frac{1}{(D+\mu)n_{max}}. \tag{B.5}$$

From eqns (B.4, B.5), given some R^* we can ensure $R > R^*$ by taking

$$R^* \leqslant \frac{c(\alpha)n_max}{\widetilde{S_d}}\left(\frac{1}{(D+\mu)n_{max}} - t(\alpha)\right)$$

$$\leqslant \frac{c(\alpha)n_max}{\widetilde{S_d}}(t_f - t(\alpha)) < R, \tag{B.6}$$

which means choosing μ and D such that

$$D + \mu \leqslant \frac{c(\alpha)}{c(\alpha)n_{max}t(\alpha) + \widetilde{S_d}\,R^*}. \tag{B.7}$$

APPENDIX C

The Second-Order Approximation

The first-order mean-field approximation was presented in Section 4.1. It was derived from the assumption that the probability of finding a site (i, j) in state S_{ij} is independent of the state of its neighbors and is uniform across the lattice. This assumption may be expressed by stating that the probability for a lattice configuration $S_{ij}(t)$ satisfies the relation:

$$P(\{S_{ij}(t)\}_{i,j\in z}) = \prod_{i,j} P(S_{i,j}(t)), \tag{C.1}$$

where the $P(S_{ij}(t))$ derive from a uniform single-site state distribution on the lattice. In Section 4.2, we also argued that the first-order approximation falls short of providing conditions for the persistence of cooperation in the system, since it cannot capture the effects of nearest-neighbors (n.n.) correlations in state.

The second-order mean-field approximation, also referred to as the Bethe–Peierls approximation (Huang, 1987), considers the state of tuple consisting of a site and its n.n. (a von Neumann environment). Such a tuple will be denoted by: $E_{i,j} \equiv (S_{i,j}, S_{i+1,j}, S_{i-1,j}, S_{i,j-1}, S_{i,j+1})$. It is assumed that the probability distribution of the tuples is independent of their environment and uniform across the lattice so that

$$P(\{S_{ij}(t)\}_{i,j\in z}) = \prod_{i,j\in z} P(E_{i,j}(t)). \tag{C.2}$$

This approximation facilitates studying n.n. correlations but ignores higher-order correlations. The class of probability distributions considered in this approximation contains the probability distributions considered in the first-order approximation as special cases, where $P(E_{i,j})$ depend only on $S_{i,j}$. Higher-order mean-field approximations consider larger tuples, thus extending the class of probability distributions further. If the system does not have long-range correlations, then as the tuples in the approximation grow it becomes more accurate and approaches the true solution of the system.

The peripheral sites within the tuple do not affect each other, and they all affect the center site in the same way. Thus, a state of a tuple can be adequately described by the state of the center site and by how many of the peripheral sites there are in each state. Therefore, a tuple E could be in one of: $3\binom{6}{2} = 45$ states. The dynamics for the probability of each state can be derived by finding the transition probabilities between states, as was done in eqns (14). A similar derivation has been used for the first-order approximation in Section 4.2. As this derivation is not very informative but is, nevertheless, incredibly tedious, we do not present it here. Results from the second-order approximation are presented in Section 4.3.

APPENDIX D

Deriving Phase Boundaries From Simulations

In this appendix, we describe how the simulated phase boundaries, from Figs 7 and 8, were derived. As noted in Section 4.1, the simplified d.d. model has three parameters that can be taken to be n_c/n_d, D/μ and $S_d = Dn_d/P_d$. We ran simulations on a 32×32 lattice, with $n_c/n_d = 2$ and D/μ varying from 1.1 to 5.0 (50 values) and Dn_d/P_d varying from 0.1 to 6.8 (50 values). The points on this parameter grid in which a non-trivial steady state was established appear in Fig. D1.

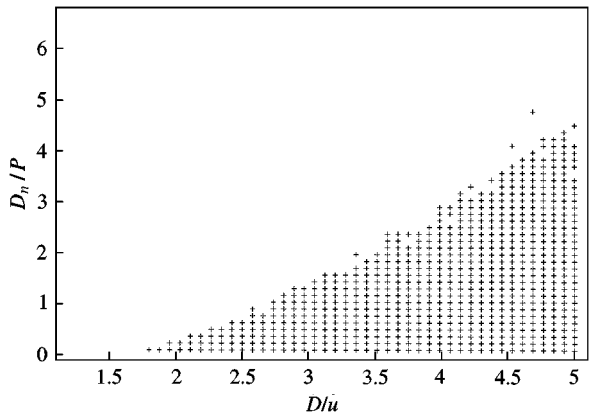

FIG. D1. Parameter values for which the simplified d.d. model reached a non-trivial steady state. The parameters taken were D/μ varying from 1.1 to 5.0 (50 values) and Dn_d/P_d varying from 0.1 to 6.8 (50 values), where for all simulations $n_c/n_d = 2$.

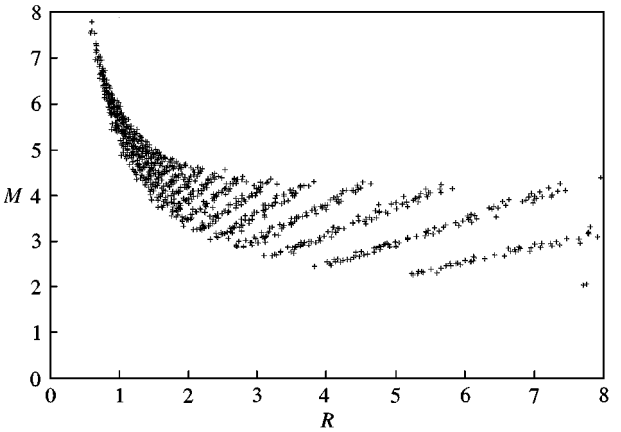

FIG. D2. The R–M phase space for the simplified d.d. model: the graph presents regions of cooperation persistence according to simulations. Simulations on a 32×32 lattice were run with the following parameters: $n_c/n_d = 2$, $P_d = 0.01$, $D/\mu = 1.1 - 5.0$ (50 values) and $S_d = Dn_d/P_d = 0.1 - 6.8$ (50 values). In each simulation in which a non-trivial steady state was established, the densities: p_e, p_c, and p_d were measured, and from them R and M were computed.

For each of the simulations that established a non-trivial steady state p_e, p_c and p_d were measured. The measurement was averaged over a long time compared to the dimension of the lattice and the typical time of a life cycle, to control the stochasticity of the simulation and the finite dimensions of the lattice. Even though we tried to control the accuracy of the measurements in individual simulations, it is not homogeneous across the parameter space. Generally, it decreases when the parameters are closer to the phase boundaries, as the time required to obtain an accurate measurement diverges at the phase boundary. From the measured p_e, p_c and p_d, we derived $R = (n_c/n_d)(p_c/p_d)$ and $M = (Dn_d/P_d)(1 + R)$. We then plotted the phase space in Fig. D2, where every point in the R–M space corresponds to a simulation that attains these values. The fact that no points are found below a certain contour, means that none of the simulations attained a steady state where such R, M values were measured. Thus, within the accuracy of the simulations, a steady state with these R, M values cannot be maintained. Based on this premise, we draw the phase boundaries in Fig. 7.

The way the phase boundaries were drawn for the d.i. model is essentially similar. The d.i. model has six parameters: α, β, γ, δ, μ and D. We ran simulations on a 32×32 lattice, with $\beta = 0.9$, $\gamma = 0$, $\delta = 1.6$, $D = 0.01$, varying $R = ((1 - \beta)/(\alpha - 1))((\alpha - \gamma)/(\delta - \beta))$ from 0.4 to 10 (50 values)

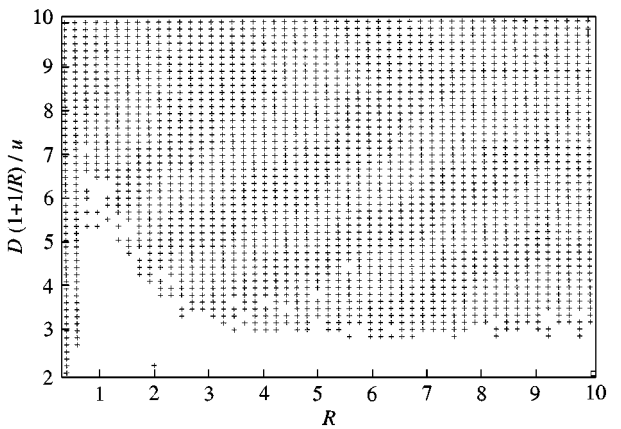

FIG. D3. Parameter values for which the simplified d.d. model reached a non-trivial steady state. The parameters taken were D/μ varying from 1.1 to 5.0 (50 values) and Dn_d/P_d varying from 0.1 to 6.8 (50 values), where for all simulations $n_c/n_d = 2$.

and $(D/\mu)(1 + 1/R)$ from 2 to 10 (50 values), where R and $(D/\mu)(1 + 1/R)$ replace parameters α and μ. The points on this parameter grid in which a non-trial steady state was established are presented in Fig. D3.

For each simulation that attained a non-trivial steady state M and R were measured (controlling for stochasticity and finite lattice size). Each simulation in which a non-trivial steady state was established appears in Fig. D4, according to the

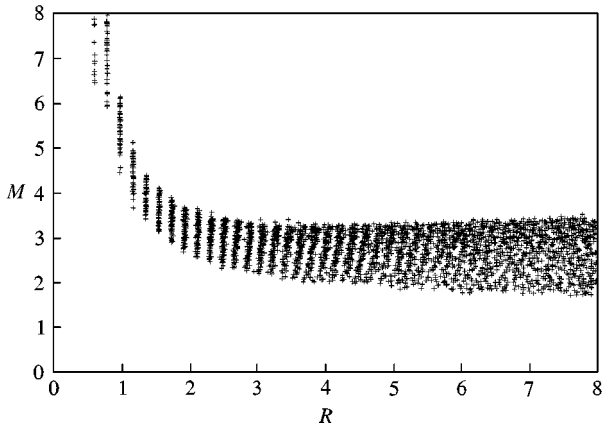

FIG. D4. The R–M phase space for the d.i. model: The graph presents regions of cooperation persistence according to simulations of the d.i. model. The simulations were done on a 32×32 lattice, using the following parameters: $\beta = 0.9$, $\gamma = 0$, $\delta = 1.6$, $D = 0.01$, $R = ((\alpha - \gamma)/(1 - \alpha))((1 - \beta)/(\delta - \beta)) = 0.4$–10 (50 values) and $(D/\mu)(1 + 1/R) = 2$–10 (50 values). Each simulation in which a non-trivial steady state was established appears in the phase space according to the R and M measured in the steady state.

measured R and M. The phase boundary in Fig. 8, was drawn using Fig. D4.

The fact that groups of points in both phase spaces appear to be on a straight line has a simple explanation. In the d.i. model, R is a parameter, and was chosen over a grid of values, whereas M is a variable deriving from the dynamics, and can therefore appear anywhere on the fixed R line. The explanation for the d.d. model is similar, $DS_d = M/(1 + 1/R)$ is a parameter, which explains the straight lines, and M and R are mutually dependent dynamical variables.

Noting that the points in Figs D2 and D4 appear to be bounded from above, we performed simulations for the d.i. model with other parameter values, to inquire whether this is in fact the case. As in Fig. D4 $\beta = 0.9$, $\gamma = 0$, $\delta = 1.6$, $D = 0.01$ and R varies from 0.4 to 10 (50 values). In one run, we took $(D/\mu)(1 + 1/R)$ to vary from 10 to 18 (50 values), and in a second run it varied between 18 and 26 (50 values). These values were chosen in such a way that they should appear above the points in Fig. D2. The points for the two new simulation series appear in Fig. D5 along with the points from the previous run, and indeed, all points established a non-trivial steady state. There appears to be no sign of an upper phase boundary.