



The Evolution of Cooperation in a Lattice-Structured Population

M. NAKAMARU[†], H. MATSUDA AND Y. IWASA

Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812-81, Japan

(Received on 11 April 1996; Accepted in revised form on 28 August 1996)

The evolution of cooperation among unrelated individuals is studied in a lattice-structured habitat, where individuals interact locally only with their neighbors. The initial population includes Tit-for-Tat (abbreviated as TFT, indicating a cooperative strategy) and All Defect (AD, a selfish strategy) distributed randomly over the lattice points. Each individual plays the iterated Prisoner's Dilemma game with its nearest neighbors, and its total pay-off determines its instantaneous mortality. After the death of an individual, the site is replaced immediately by a copy of a randomly chosen neighbor. Mathematical analyses based on mean-field approximation, pair approximation, and computer simulation are applied. Models on one and two-dimensional regular square lattices are examined and compared with the complete mixing model. Results are: (1) In the one-dimensional model, TFT players come to form tight clusters. As the probability of iteration w increases, TFTs become more likely to spread. The condition for TFT to increase is predicted accurately by pair approximation but not by mean-field approximation. (2) If w is sufficiently large, TFT can invade and spread in an AD population, which is impossible in the complete mixing model where AD is always ESS. This is also confirmed by the invasion probability analysis. (3) The two-dimensional lattice model behaves somewhat in between the one-dimensional model and the complete mixing model. (4) The spatial structure modifies the condition for the evolution of cooperation in two different ways: it facilitates the evolution of cooperation due to spontaneously formed positive correlation between neighbors, but it also inhibits cooperation because of the advantage of being spiteful by killing neighbors and then replacing them.

© 1997 Academic Press Limited

Introduction

Human society includes many forms of cooperation between unrelated individuals: some are explicitly institutional such as social welfare but many others are less formal. A key to the success of a person in human societies is the ability to make trustful cooperative relationships with many other people. In the modern society especially, cooperation is not restricted to relatives, although the role of cooperation amongst kin is often more important in traditional societies (Alexander, 1979).

Non-human animals also cooperate, but the cooperation among relatives is used, as exemplified by sterile workers of social insects (Hamilton, 1964). Cooperation among non-relatives can also evolve,

however, through "reciprocal altruism" (Trivers, 1971). For instance, vampire bats share food with individuals who in turn also share food (Wilkinson, 1984). Cooperation among relatives and non-relatives in common and pigmy chimpanzees (*Pan troglodytes* and *P. paniscus*) is even more elaborate (de Waal, 1982).

A Prisoner's Dilemma game clearly illustrates the difficulty of maintaining cooperation between players in spite of its advantage. The game is played by a pair of individuals, who have two options: either to cooperate (C) or to defect (D). If both players cooperate, both get pay-off R , standing for "reward". If one defects while the other cooperates, the one who plays C gets a pay-off S , standing for "sucker", while the one who plays D gets a pay-off of T , standing for "temptation to defect". If both defect, both get pay-off P , standing for "punishment" (see Table 1 for pay-offs). The order of the magnitude of pay-offs

[†] Author to whom correspondence should be addressed.
E-mail: nakamaru@mbox.biology.kyushu-u.ac.jp

TABLE 1
The pay-offs of the Prisoner's Dilemma game for player A to obtain by interacting with player B

		Player B	
		Cooperate	Defect
Player A	Cooperate	$R = 3$	$S = 0$
	Defect	$T = 5$	$P = 1$

These values are used in computer simulations. See text for explanation of each element.

is $S < P < R < T$. In addition, we assume that the pay-offs satisfy $2R > T + S$, which prevents a strategy of "Alternation of D and C" from invading in a cooperative population. Although both players receive a higher pay-off if they cooperate than if they defect ($R > P$), cooperation is difficult to maintain because each player would have a higher pay-off by defection than by cooperation irrespective of the partner's action ($T > R$, and $P > S$). This causes a dilemma—each player's attempt to increase its own pay-off results in a smaller pay-off (P) because of the failure to cooperate.

One way to resolve the dilemma is the possibility that the same pair of contestants may play more than once and that the total number of times the game is played between the same pair is unpredictable. Axelrod & Hamilton (1981) and Axelrod (1984) studied the evolution of cooperation using an iterated Prisoner's Dilemma game. They collected various potential strategies from all over the world and then let them play the Prisoner's Dilemma game with each other in a computer. The strategy with the highest score in this computer tournament was Tit-for-Tat (abbreviated as TFT), which cooperates on the first move and then does whatever the other player did on its previous move. The final population should be evolutionarily stable, implying that no mutant strategy can invade it. This result can be regarded as supporting the evolution of cooperation under iteration in which defection is the only evolutionarily stable strategy without iteration. The cooperative strategy TFT is more likely to be evolutionarily stable if w , the probability of playing the same partner again, is larger.

Axelrod (1984) studied the effect of spatial structure in interaction and reproduction (he called it "territoriality") on the evolution of cooperation. If interactions among individuals occur within an area much smaller than the whole population, and if the individuals of the same strategy tend to form tight clusters, then a cooperative strategy would be more likely to spread and be maintained in a spatially structured than in a perfectly mixed population.

Neighbors tend to become occupied by players of the same strategies if migration is limited in spatial range or if players tend to imitate their neighbors. To study the effect of spatial structure, Axelrod (1984) used a lattice model in which each lattice point is occupied by a single individual which plays the game only with its nearest neighbors. He carried out a computer simulation of the spatial patterns when one TFT (a cooperative strategy) mutated to an All-Defect (abbreviated as AD, a strategy which defects on every move) in a two-dimensional spatial lattice occupied by TFT only. Each individual repeatedly played the iterated Prisoner's Dilemma game with its nearest neighbors. The average pay-offs gained by these interactions were calculated, and then each individual player changed its strategy to the one adopted by the most successful nearest neighbor who had achieved the highest score.

Nowak & May (1992) studied a Prisoner's Dilemma game on a similar lattice structured population. To concentrate on the effect of spatial structure, rather than the effect of iteration in enhancing the evolution of cooperation, they simplified the game's structure. For example, they used the strategy of all cooperate, AC, instead of TFT, which implies that the game is played once only in each time we evaluate the score ($w = 0$). They also assumed specific values for the pay-offs: $R = 1$, $T = b > 1$, $S = P = 0$. They observed that, for a limited range of b , AC and AD can coexist and produce a complex and constantly changing spatial patterns.

There are theoretical studies of the evolution of altruistic social behavior which do not mention the Prisoner's Dilemma game. For example, Matsuda (1987) and Matsuda *et al.* (1987) studied the evolution of cooperation on a lattice structured model, in which two types differing in their social interaction compete with each other. Assuming that some sites are vacant and that reproduction occurs only to nearest neighbor vacant sites, each lattice site takes one of the three alternative states. The model is a continuous time Markov chain, in which the state transition is not synchronized as in Axelrod (1984) or in Nowak & May (1992). Matsuda (1987) and Matsuda *et al.* (1987) discovered that the effect of the spatial clumping to the relative advantage of cooperation changes with the density. A similar conclusion was derived by Taylor (1992) and Wilson *et al.* (1992), who concluded that population viscosity (low mobility and a locally limited interaction) is not very effective in promoting the evolution of altruism. The major reason for this result is that the advantage of an enhanced probability for altruists

to be surrounded by other altruists in a viscous population is cancelled out by the disadvantage of the lack of surrounding vacant sites needed for reproduction.

Recently, lattice structured models have been used in describing ecological processes of sessile organisms, such as terrestrial plants or marine benthic invertebrates. The models have been studied mostly by computer simulation of the spatial stochastic processes. Sometimes, the results are compared with dynamics derived by the neglect of spatial correlation, or by assuming perfect mixing (mean-field approximation) (Caswell & Etter, 1992; Durrett & Levin, 1994a, b). However, direct computer simulation of a stochastic lattice model is very costly in computation time, and in general it is often difficult to gain insight into a model's behavior only from computer simulations. In addition, the effect of spatial structure sometimes produces predictions qualitatively different from non-structured populations (e.g. Harada & Iwasa, 1994; Sato *et al.*, 1994). Matsuda *et al.* (1992) developed a method to construct a closed dynamical system of overall densities and correlation between nearest neighbors, by adopting pair approximation for lattice structured population dynamics. This approach constructs a system of ordinary differential equations for the average population densities and the local densities, the latter giving the nearest neighbor correlation of states.

In this paper, we study both one-dimensional and two-dimensional lattice models where individuals using either TFT or AD play the iterated Prisoner's Dilemma game with its neighbors. The accumulated pay-offs then determine the mortality. After the death of an individual, the site is occupied by a copy of a randomly chosen neighbor [the nearest-neighbor migration model of Matsuda (1987)]. The model is a continuous-time Markov chain. Although we call the transition of state the "death" of an individual followed by "reproduction" throughout the paper, it may in fact also represent an individual changing its strategy to adopt a new one randomly sampled from its neighbors, which would be more appropriate for human societies.

In addition to direct computer simulations, we derive dynamics based on mean-field approximation and those based on pair approximation which traces both the total frequency of the two strategies and the nearest neighbor correlation. Based on these results, we show that the spatially limited interaction may facilitate or inhibit the evolution of cooperation.

TABLE 2
The expected total pay-offs obtained by a player on site A through interacting with a player on site B

		Site B	
		TFT	AD
Site A	TFT	$R/(1-w)$	$S + wP/(1-w)$
	AD	$T + wP/(1-w)$	$P/(1-w)$

Iterated Prisoner's Dilemma in a Completely Mixed Population

First we consider an iterated Prisoner's Dilemma game. Two players interact repeatedly, and may change their action depending on the partner's previous act. In most of the computations in this paper, we used the following parameter values for the pay-offs in a single interaction: $R = 3$, $T = 5$, $S = 0$, and $P = 1$ (see Table 1). We consider two strategies for each player to adopt: One strategy is Tit-for-Tat (TFT), which first cooperates, then adopts the same action as what the partner did previously. The second strategy is All-Defect (AD), which always defects. Let w be the probability that the same two players interact in the following step as well, and $w^{n-1}(1-w)$ be the probability that they interact exactly n times ($n = 1, 2, 3, \dots$). The expected number of times the two players interact is $1/(1-w)$.

When a TFT plays with an AD, the TFT gets pay-off S in the first round. In the second round TFT gets P with the probability that they interact again, w . In the third round TFT gets P with the probability that they interact at least three times, w^2 and so on. Then the expected score of TFT whose partner is AD is $V(T/D) = S + Pw + Pw^2 + Pw^3 + \dots = S + wP/(1-w)$. Let $V(i/j)$ be the expected score of a player with strategy i obtained from the interaction with a neighbor adopting strategy j ($i, j = T$ or D , indicating TFT and AD, respectively). Table 2 shows the expected total pay-offs $V(i/j)$.

Before examining the game on a lattice, we first study the complete mixing model, in which each player interacts with another player randomly chosen from the whole population and the reproductive rate is dependent on the score but is independent of the availability of vacant sites in the neighborhood.

If the total population size is sufficiently large, the fitnesses of TFT and AD players, denoted by F_T and F_D , are expressed by using ρ_T and ρ_D , the fractions of TFT and AD players, respectively in the population (note $\rho_T + \rho_D = 1$, as there is no vacancy):

$$F_T = F_0 + \frac{R}{1-w} \rho_T + \left(S + \frac{wP}{1-w} \right) \rho_D \quad (1a)$$

$$F_D = F_0 + \left(T + \frac{wP}{1-w} \right) \rho_T + \frac{P}{1-w} \rho_D \quad (1b)$$

where F_0 is the baseline fitness that a player enjoys without social interaction. Other terms are the average pay-offs in Table 2 with the weight of the fraction of two strategies in the population. Then the density of TFT increases if $F_T > F_D$. This inequality is rewritten as:

$$\rho_T > \frac{(1-w)(P-S)}{(S+T-2P)w + R - S - T + P} = \frac{1-w}{3w-1}. \quad (2)$$

The shadow area of Fig. 1 illustrates condition (2). The horizontal axis is for w . The range of w for TFT to be evolutionarily stable is given by eqn (2) with $\rho_T = 1$. It is rewritten as $w > (T-R)/(T-P) = 0.5$. In contrast, eqn (2) always fails if $\rho_T = 0$ and $w < 1$, implying that TFT if rare cannot invade a population dominated by AD. Hence AD is evolutionarily stable when $w < 1$ (see Axelrod & Hamilton, 1981).

In addition, Axelrod & Hamilton proved that a third strategy ‘‘Alternation of D and C’’ cannot invade the TFT population if $w \geq (T-R)/(R-S)$. Axelrod [1984; Proposition (ii) in Appendix B] hence concluded that TFT is collectively stable if and only if w is larger than both $(T-R)/(T-P)$ and $(T-R)/(R-S)$ [however, see Boyd & Lorberbaum (1987) and Boyd (1989)].

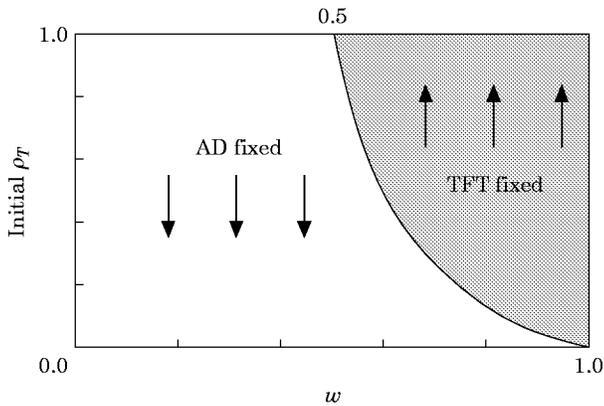


FIG. 1. The phase plane of completely mixed model, the evolutionary game with fitness given by eqn (1). Horizontal axis is the probability of reiteration w , vertical axis is the density of TFT ρ_T . The region in which ρ_T increases is indicated by shade, given by eqn (2). For w less than 0.5, TFT should always decrease with time and AD is the only evolutionarily stable strategy. In contrast, for $w > 0.5$, TFT can be stable as it refuses the invasion of rare AD. However, AD is always evolutionarily stable because rare TFT cannot invade into a population dominated by AD.

One-dimensional Lattice Model

Now, we study the iterated Prisoner’s Dilemma on a lattice structured habitat. In the initial population, each lattice site is filled either by TFT or AD randomly with a given probability. Then each player engages in the iterated Prisoner’s Dilemma game with their neighbors. Let z be the number of neighbors with whom a player interacts. In a one-dimensional lattice, $z = 2$. In the two-dimensional lattice, each player has eight neighboring sites ($z = 8$). The mortality of the player is determined by its total score, which is the sum of the scores obtained by interacting with the z neighbors. After the death of an individual, the empty lattice site is filled immediately by a copy of a neighbor randomly chosen among z possible sites. This is called the nearest neighbor migration model by Matsuda (1987). The model is a continuous time Markov chain, and in a sufficiently short time interval only a single event of state transition occurs.

The total score B of an individual depends on the strategy adopted (TFT or AD) and on the number of TFT neighbors, denoted by n ($0 \leq n \leq z$). Let $B_{T,n}$ and $B_{D,n}$ be the scores of a TFT player and an AD player, respectively, if surrounded by n TFT neighbors and $(z-n)$ AD players. They are:

$$B_{T,n} = nV(T/T) + (z-n)V(T/D) \quad (3a)$$

and

$$B_{D,n} = nV(D/T) + (z-n)V(D/D). \quad (3b)$$

The mortality of a player is a decreasing function of the total score. Specifically we assume that the instantaneous mortality of a TFT player surrounded by n TFT and $(z-n)$ AD decreases exponentially with B :

$$M_{T,n} = \exp(-\alpha \times B_{T,n}) \quad (4a)$$

and that of an AD player is:

$$M_{D,n} = \exp(-\alpha \times B_{D,n}). \quad (4b)$$

In a short time interval of length Δt , a player dies with probability $M_{T,n} \Delta t$ or $M_{D,n} \Delta t$. Constant α was chosen so that the mortality varies between sites of different states and a different combination of neighbors. For a reiteration probability w close to one, the total score becomes very large, because the expected number of iterations is proportional to $1/(1-w)$, and the mortality of different types in eqn (4b) are all very small. To prevent this situation, we chose α as a decreasing function of w , specifically $\alpha = 0.2(1-w)$ in this paper.

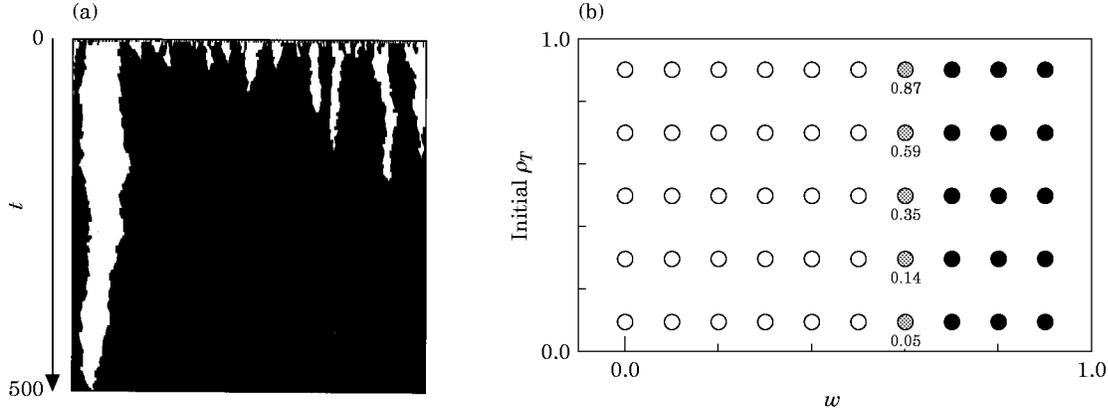


FIG. 2. Computer simulations of one-dimensional lattice model. (a) The spatial patterns generated by the model when $w = 0.8$ and initial ρ_T is 0.5. There are 500 sites, and black and white points are for those occupied by TFT and those occupied by AD, respectively. Vertical axis is for time. The initial population ($t = 0$) is random (as shown by the top row), but a strongly clumped spatial distribution is quickly formed, in which the lattice is composed of long runs of TFT and those of AD. Finally TFT won in this simulation. (b) The outcome of simulation for different iteration probability w and initial density ρ_T . For each pair of parameters, 100 runs were computed. Open circles indicate sets of parameters for which all 100 runs end up with the extinction of TFT and the fixation of AD, while solid circles are parameter values for which all the runs show the fixation of TFT (the fractions are indicated by numerals), but others with the fixation of AD. When w is less than 0.6, AD is an evolutionarily stable strategy. When 0.6 is larger than 0.6, TFT is evolutionarily stable.

COMPUTER SIMULATION OF A ONE-DIMENSIONAL LATTICE MODEL

We carried out computer simulation of the model, on a one-dimensional lattice ($z = 2$). To remove the effect of edges, we used a periodic boundary condition: i.e. the lattice is circular (the rightmost site is the nearest neighbor of the leftmost site), with a lattice size of 500. In the initial pattern sites were filled independently either by TFT or AD with a given probability. We computed the model until a time in which either one of the two types occupied the whole population. We computed runs with the initial population with different fractions of TFT: ρ_T with 0.1, 0.3, 0.5, 0.7, and 0.9.

Examples of spatial patterns of the model are in Fig. 2(a). Starting from a random spatial pattern, clusters composed of the same type are quickly formed. Let ρ_T be the total fraction of TFT in the whole lattice, called the global density of TFT, and ρ_D be the global density of AD. Figure 2(b) illustrates the final outcome of the simulation for different initial density ρ_T and parameter w . When w is less than 0.6, AD is finally fixed in the population for any initial global density ρ_T . When the parameter w is larger than 0.6, TFT became fixed for any initial density ρ_T . No evolutionary bistability was observed.

These results are quite different from the behavior of the complete mixing model (Fig. 1). Next we apply analytical calculations to confirm the computer simulation in Fig. 2(b).

DYNAMICS OF DENSITY

The time change of ρ_T , the global density of TFT, is given by:

$$\begin{aligned} \frac{d\rho_T}{dt} = & -M_{T,1}[\text{density of DTT}] \\ & -M_{T,0}[\text{density of DTD}] \\ & +M_{D,1}[\text{density of DDT}] \\ & +M_{D,2}[\text{density of TDT}]. \end{aligned} \quad (5)$$

Here we call a site occupied by a TFT player as a T-site, and a site occupied by an AD player as a D-site for short. The first term on the r.h.s. of eqn (5) indicates the rate at which a T-site located between a D-site and a T-site changes to a D-site. It is a product of the mortality of T, denoted by $M_{T,1}$, and the density of the triplet ‘‘DTT’’ in the lattice, i.e. the probability of a randomly chosen triplet is ‘‘DTT’’. We need to consider a factor 2 because there is another triplet ‘‘TTD’’ having the same contribution as ‘‘DTT’’. However, this factor is canceled out by another factor 1/2, the probability for the middle T to be replaced by D instead of T. Similarly the second term on the r.h.s. is for the transition of T in a triplet ‘‘DTD’’ to D. The third and the fourth terms are for the transition of D to T in the middle of ‘‘DDT’’ and in the middle of ‘‘TDT’’, respectively.

The frequency of a triplet ‘‘DTT’’ cannot be expressed by only using global densities, such as ρ_T and ρ_D . We need to introduce a conditional probability for a site to be T if it is chosen next to a

T site. This is expressed as $q_{T|T}$ and called conditional density or local density of T sites (Matsuda *et al.*, 1992; Harada & Iwasa, 1994). In a similar way, $q_{a|bc}$ $\{\alpha, b, c = T \text{ or } D\}$ indicates a conditional density of a higher order. For example, $q_{T|TD}$ is the probability that TFT is in the neighborhood of TFT whose neighbor is AD. Using these notations, the density of triplet DTT is $\rho_T q_{T|T} q_{D|TT}$. We can rewrite the frequencies of the various triplets in eqn (5) as:

$$\begin{aligned} \frac{d\rho_T}{dt} = & -M_{T,1}\rho_T q_{T|T} q_{D|TT} - M_{T,0}\rho_T q_{D|T} q_{D|TD} \\ & + M_{D,1}\rho_D q_{T|D} q_{D|DT} + M_{D,2}\rho_D q_{T|D} q_{T|DT}. \end{aligned} \quad (6)$$

Hence, to calculate the dynamics of the average density of TFT, we need to know conditional densities, such as $q_{T|T}$ and $q_{T|TT}$, which include information concerning the correlation of states between close neighbors.

In the following sections, two methods of constructing a closed dynamical system, based on ‘‘mean-field approximation’’ and ‘‘pair approximation’’ are developed.

MEAN-FIELD APPROXIMATION

Mean-field approximation is to neglect spatial structure or to assume a random spatial configuration. Under this assumption, the local density and other probability on the interaction between more than two players are the same as the global density: $q_{T|ab} = q_{T|c} = \rho_T \{\alpha, b, c = T \text{ or } D\}$. Then the density of triplet DTT is simply $\rho_T^2 \rho_D$. If we adopt this simplification and $\rho_D = 1 - \rho_T$, the dynamics of global density given by eqn (6) become:

$$\begin{aligned} \frac{d\rho_T}{dt} = & \rho_T(1 - \rho_T) \{ -M_{T,1}\rho_T - M_{T,0}(1 - \rho_T) \\ & + M_{D,1}(1 - \rho_T) + M_{D,2}\rho_T \}. \end{aligned} \quad (7)$$

These dynamics of a single variable have two trivial equilibria: $\rho_T = 0$ and $\rho_T = 1$. In addition, there may be an intermediate equilibrium with $0 < \rho_T < 1$, such as:

$$\begin{aligned} \rho_T = & \frac{M_{T,0} - M_{D,1}}{M_{T,0} - M_{D,1} - M_{T,1} + M_{D,2}} \\ = & \frac{e^{\alpha(T+P-2S)} - 1}{e^{\alpha(T+P-2S)} - 1 - e^{\alpha\left((T-S) - \frac{R-P}{1-w}\right)} + e^{-\alpha(T-P)}}. \end{aligned} \quad (8)$$

The intermediate equilibrium (8) is feasible only when $w > w_a = (2T - S - (R/2T) - S - P)$, which is $w_a = \frac{7}{9} = 0.778$ for pay-offs in Table 1.

By examining the sign of eqn (7), we know that, if $0 \leq w \leq w_a$, $\rho_T = 1$ is unstable and $\rho_T = 0$ is globally stable. The extinction of TFT is inevitable, and AD is the unique evolutionarily stable strategy. In contrast, if $w_a < w < 1$, both $\rho_T = 1$ and $\rho_T = 0$ are locally stable, and the intermediate equilibrium given by eqn (8) is unstable. The system is bistable and the evolutionary end point depends on the initial density of TFT. If the density of TFT players is higher than the unstable point, TFT players tend to increase and become fixed. If not, AD players occupy the whole population instead. Hence TFT is evolutionarily stable when reiteration probability w is larger than w_a , and AD is always evolutionarily stable.

Figure 3(a) illustrates the change in ρ_T for different w . The phase plane is separated to two regions again, and the pattern is unable to explain the results of computer simulation [Fig. 2(b)]. First, the threshold value of w for a TFT population to be locally stable from computer simulation is about 0.6, clearly smaller than $\frac{7}{9} (= 0.778)$ predicted by the mean-field approximation. Second, the system does not show bistability for any value of w —either a TFT population or an AD population is globally stable depending on w . Especially notable is that the shaded area in which TFT is predicted to increase is much smaller in Fig. 3(a) than in Fig. 1. This implies that cooperation is much more unlikely to evolve in the dynamics of a mean-field approximation of the model than in the complete mixing model, in spite of the fact that the spatial configuration of individuals are neglected in both models. We will discuss the reason for this difference later.

PAIR APPROXIMATION

Pair approximation is a method of constructing a system of ordinary differential equations for the global and local densities. In some cases, it predicts the population dynamics of lattice structured models accurately, while the equations based on the mean-field assumption, such as eqn (7), fail to do so (Sato *et al.*, 1994; Harada *et al.*, 1995). Pair approximation keeps the distinction between the average density and the local density, but does not consider the correlation beyond nearest neighbors.

Note that pair approximation does not assume perfect independence between nonnearest neighbor sites. It assumes that the correlation between nonnearest neighbor sites can be approximated by the product of the nearest neighbor correlation. For example, consider a triplet, ABC, in which A and B are correlated in state and B and C are correlated because they are neighbors. Then A and C are also correlated. Pair approximation assumes that the

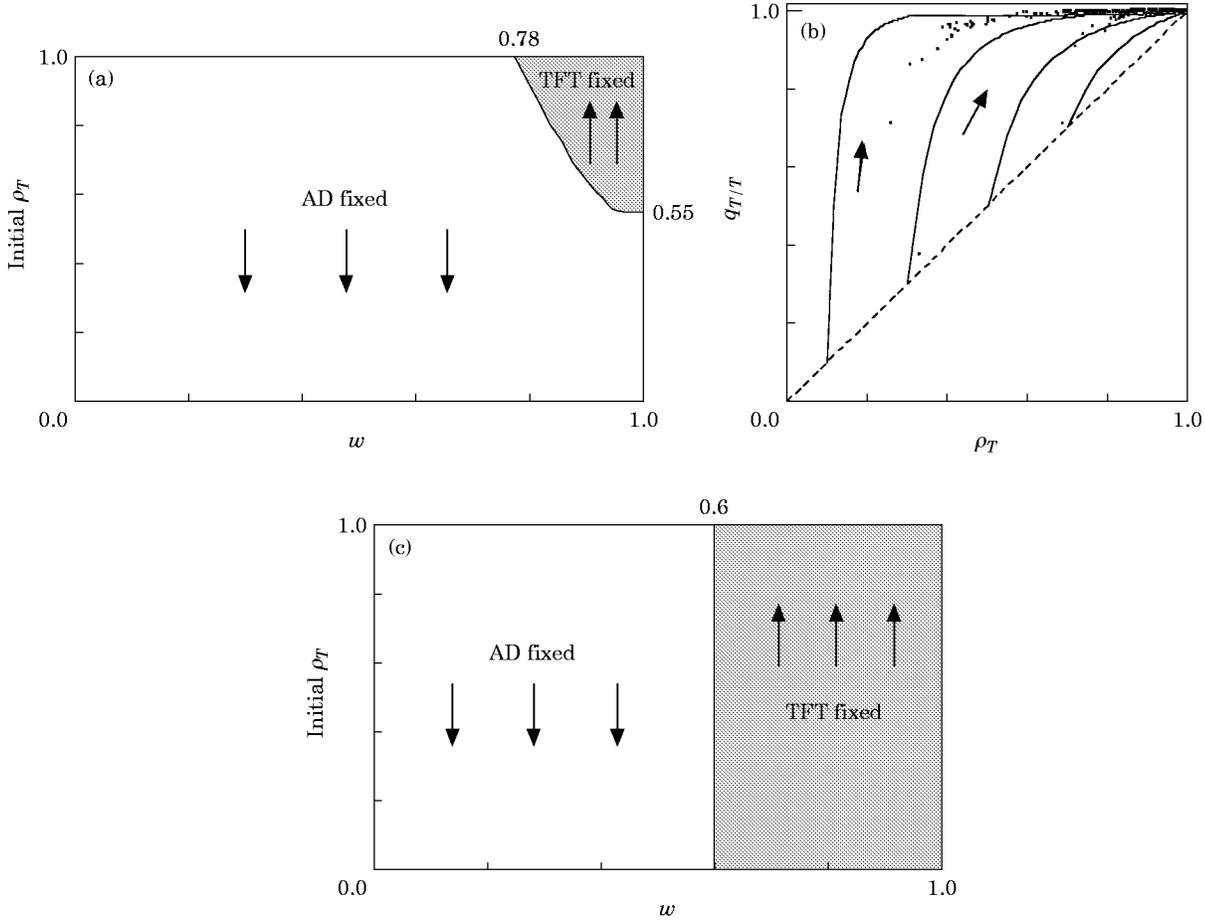


FIG. 3. Analyses of one-dimensional lattice model: (a) The phase plane of the dynamics based on mean-field approximation, eqn (7). Horizontal axis is reiteration probability w , vertical axis is the density of TFT ρ_T . The region in which ρ_T increases is indicated by shade. Note that this region is much smaller than that for the complete mixing case (Fig. 1), and the area indicated by computer simulation [circles in Fig. 2(b)]. This implies that mean-field dynamics predict a much more stringent condition for the evolution of cooperation than observed by simulation or predicted by the model without spatial structure. AD is always evolutionarily stable. The system shows bistability when w is larger than w_c . (b) Curves are the trajectories of the dynamics of pair approximation when w is 0.8. Dots are from computer simulation where the time interval between points is 10. A broken line is for $q_{T|T} = \rho_T$ implying a random spatial pattern. The horizontal axis is global density ρ_T and the vertical axis is local density $q_{T|T}$. Arrows indicate the direction of movement along trajectories. $q_{T|T}$ keeps increasing and hence TFT clusters become larger and larger. (c) The phase plane of the dynamics based on pair approximation, eqns (12). Horizontal axis is the probability of reiteration w , vertical axis is the density of TFT ρ_T . The region in which ρ_T increases is indicated by shade. When w is less than w_b , AD is an evolutionarily stable strategy. When w is larger than w_b , TFT is an evolutionarily stable strategy and the AD is no longer evolutionarily stable. The predictions by the dynamics based on pair approximation are consistent with the results of computer simulation of Fig. 2(b).

correlation between A and C is caused only by the indirect effect via their common neighbor B. In reality, A and C can be more or less strongly correlated than the simple product of correlation coefficients between the two nearest neighbors, and hence the pair approximation may not be exact.

Let ρ_{TT} be the doublet density, i.e. the probability that a randomly chosen pair of nearest neighbors are both TFT. Then the local density of TFT is the ratio, $q_{T|T} = \rho_{TT}/\rho_T$. Hence we have:

$$\frac{dq_{T|T}}{dt} = \frac{1}{\rho_T} \frac{d\rho_{TT}}{dt} - \frac{q_{T|T}}{\rho_T} \frac{d\rho_T}{dt}$$

which implies that the dynamics of doublet density are needed to compute the dynamics of local density. The time change is ρ_{TT} is derived as follows:

$$\begin{aligned} \frac{d\rho_{TT}}{dt} = & -M_{T,1}\rho_T q_{T|T} q_{D|TT} \\ & + M_{D,1}\rho_D q_{T|D} q_{D|DT} + 2M_{D,2}\rho_D q_{T|D} q_{T|DT}. \end{aligned} \quad (9)$$

The first term of the r.h.s. of eqn (9) indicates the rate of transition from a triplet TTD to another triplet TDD. The second term is the rate of transition from DDT to DTT, which produces a new TT pair. Then the last term is the rate of transition from TDT to

TTT, which creates two new ‘‘TT’’ pairs expressed by factor 2.

To construct a closed dynamical system of global density ρ_T and local density $q_{T|T}$, we neglect the correlation of state beyond nearest neighbor pairs:

$$q_{abc} = q_{ab}\{a, b, c = T \text{ or } D\} \quad (10)$$

which is called pair approximation, or doublet decoupling approximation (Matsuda *et al.*, 1992; Sato *et al.*, 1994). Note also $\rho_{TD} = \rho_{DT}$. In addition, we have the following relations, coming from the definition of conditional probabilities:

$$\rho_D = 1 - \rho_T \quad (11a)$$

$$q_{D|T} = 1 - q_{T|T} \quad (11b)$$

$$q_{D|D} = \frac{1 - 2\rho_T + \rho_T q_{T|T}}{1 - \rho_T} \quad (11c)$$

$$q_{T|D} = \frac{\rho_T(1 - q_{T|T})}{1 - \rho_T}. \quad (11d)$$

Then a pair of ordinary differential equations for global density and local density are derived:

$$\begin{aligned} \frac{d\rho_T}{dt} = & \rho_T(1 - q_{T|T}) \left\{ q_{T|T} \left(-M_{T,1} \right. \right. \\ & + M_{T,0} + \frac{M_{D,1}\rho_T}{1 - \rho_T} - \frac{M_{D,2}\rho_T}{1 - \rho_T} \left. \right) - M_{T,0} \\ & \left. + M_{D,1} \frac{1 - 2\rho_T}{1 - \rho_T} + \frac{M_{D,2}}{1 - \rho_T} \right\} \quad (12a) \end{aligned}$$

$$\begin{aligned} \frac{dq_{T|T}}{dt} = & (1 - q_{T|T})^2 \left\{ q_{T|T} \left(-M_{T,1} \right. \right. \\ & + M_{T,0} + \frac{M_{D,1}\rho_T}{1 - \rho_T} - \frac{M_{D,2}\rho_T}{1 - \rho_T} \left. \right) \\ & \left. + M_{D,1} \frac{1 - 2\rho_T}{1 - \rho_T} + \frac{2M_{D,2}}{1 - \rho_T} \right\}. \quad (12b) \end{aligned}$$

This system of autonomous equations can be analysed by the standard technique for nonlinear dynamics. All the points on line $q_{T|T} = 1$ are equilibria. The asymptotic behavior of the system is different depending on whether the reiteration probability w is larger than a critical value $w_b = (T - S - R + P)/(T - S) = 0.6$. The trajectories converge to $(\rho_T, q_{T|T}) = (0, 1)$ for $w < w_b$; they converge to $(\rho_T, q_{T|T}) = (1, 1)$ for $w > w_b$. When

$w = w_b$, different trajectories converge to different points on the line of equilibria $q_{T|T} = 1$, which are neutrally stable. Figure 3(b) shows the trajectories of pair approximation dynamics, eqns (12), when $w = 0.8$.

Figure 3(c) is the phase plane for pair-approximation dynamics, eqn (12). Then when w is smaller than w_b , which is 0.6 for pay-offs in Table 1, AD is evolutionarily stable, and when w is larger than w_b TFT is evolutionarily stable. No bistability is predicted. These are consistent with the results of computer simulation [Fig. 2(b)].

Invasion Success Probability

Although a cooperative behavior is evolutionarily stable in the complete mixing model, it is quite difficult for the cooperation to establish itself in a population initially dominated by non-cooperative behavior, because AD is always evolutionarily stable and repels the invasion of rare TFT. However, in the one-dimensional lattice model, both computer simulation and pair-approximation dynamics show that TFT can invade successfully a population dominated by AD if the probability of iteration w is sufficiently high.

Computer simulation shows that TFT and AD quickly form clusters of the same strategy and that the movement of their boundaries determines the fate of the system [Fig. 2(a)]. Changes of spatial patterns occurs if TFT in the edge of a TFT cluster becomes AD or if AD in the edge of the AD cluster changes to TFT. The location of the boundary between two clusters follows a Random-Walk. Then the condition for the successful invasion of TFT in the population occupied by AD players can be derived by the standard techniques for stochastic processes.

In Appendix A, we derive the probability $Q_T = 1 - Q_1$ that TFT survive ultimately starting from a single TFT player as eqn (A.8). It is possible for $w > w_b$ where $w_b = (T - S - R + P)/(T - S)$ is the critical value, which is $w_b = 0.6$ if pay-offs in Table 1 are used. In contrast, $Q_T = 0$ for $w \leq w_b$.

Similarly, the probability for a single AD player to invade a population composed of TFT can be derived by the equations based on birth-and-death processes. The probability that AD survives in the lattice where only one AD existed in the initial population, denoted by $R_D = 1 - R_1$, is given by eqn (A.11). R_D is positive for $w < w_b$, but zero for $w \geq w_b$.

The two lines in Fig. 4 illustrate the probabilities of successful invasion for different w . We see that TFT is evolutionarily stable for w larger than w_b and that AD is the evolutionarily stable strategy for w less than

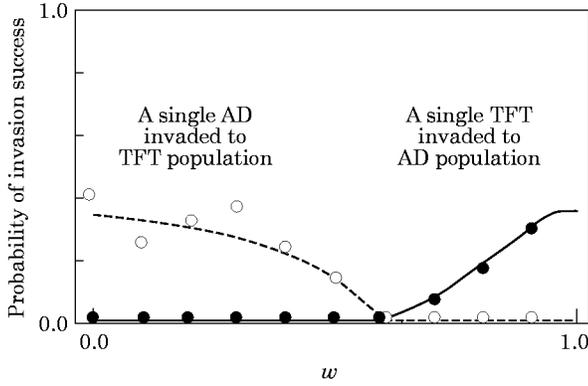


FIG. 4. The probability of invasion success. A solid and broken lines are the predictions by eqns (A.8) and (A.11) derived by the birth-and-death processes. Solid and open circles show the results of the computer simulations, which are consistent with the predictions. A single TFT invader in the population occupied by AD can increase in number and becomes fixed with a positive probability if $w > 0.6$ but must go extinct if $w < 0.6$. In contrast, a single AD invader in the population occupied by TFT can increase and becomes fixed with a positive probability if $w < 0.6$ but cannot increase if $w > 0.6$. These results are consistent with the predictions by pair-approximation [Fig. 3(c)].

w_b . Computer simulation results using a finite sized lattice (500 sites) are indicated by solid and open circles, and are consistent with the curves given by eqns (A.8) and (A.11).

If the invasion of a single mutant is not a unique event but occurs recurrently, then a small but positive probability of establishment is sufficient for the ultimate success of invasion. Figure 4 therefore implies that TFT beats AD for $w > w_b$ and AD beats TFT for $w < w_b$. This is consistent with the results of pair-approximation [Fig. 3(c)].

Two-dimensional Lattice

We also examined the model on a two-dimensional square lattice. The number of nearest neighbors with whom a single player interacts is $z = 8$ (i.e. we assume Moore neighborhood instead of Neumann neighborhood of $z = 4$).

Just as for eqn (6) for the one-dimensional model, the time change of global density can be expressed as:

$$\frac{d\rho_T}{dt} = - \sum_{n=0}^{z-1} \binom{z}{n} M_{T,n} [\text{density of } \{T, n\}] \frac{z-n}{z} + \sum_{n=1}^z \binom{z}{n} M_{D,n} [\text{density of } \{D, n\}] \frac{n}{z} \quad (13a)$$

where $\{T, n\}$ indicates a T-site surrounded by n

T-sites and $(z - n)$ D-sites. With pair approximation, it can be expressed as:

$$[\text{density of } \{T, n\}] \approx \rho_T (q_{T|T})^n (q_{D|T})^{z-n}. \quad (13b)$$

The first term in eqn (13a) indicates the transition of a TFT to an AD given that it interacts with n TFT and $(z - n)$ AD. The second term comes from the transition of an AD to a TFT if it interacts with n TFT and $(z - n)$ TFT.

MEAN-FIELD APPROXIMATION

If mean-field approximation is adopted, eqns (13) are rewritten as eqn (B.2) in Appendix B. This equation has two trivial equilibria: $\rho_T = 0$ and $\rho_T = 1$. In addition, there may be an intermediate equilibrium ($0 < \rho_T < 1$) if $w > w_d = (40 - S - 7T)/(32 - S + P)$, which is $w_d = 19/33 = 0.576$ for the pay-offs in Table 1.

The phase plane of w vs ρ_T is given in Fig. 5(a). This result suggests that AD is always the evolutionarily stable strategy (when $w < 1$). When $w > w_d$ TFT is also evolutionarily stable, and the system is bistable.

PAIR APPROXIMATION

By adopting pair approximation, we can construct a closed dynamical system of ρ_T and $q_{T|T}$, as eqns (B.4) in Appendix B. The system of autonomous equations such as eqns (B.4) can be analysed by the standard techniques of nonlinear dynamics. All the points on line $q_{T|T} = 1$ are equilibria. The asymptotic behavior of the system depends on the value of w .

According to numerical analysis, all the trajectories starting from internal points converge to $\rho_T = 0$ for $0 \leq w \leq 0.49$, and all the trajectories converge to $(\rho_T, q_{T|T}) = (1, 1)$ for $w \geq 0.77$. For w between 0.77 and 0.49, whether the trajectories converge to $\rho_T = 0$ or to $(\rho_T, q_{T|T}) = (1, 1)$ depends on the initial density of TFT. The system is bistable. For example, $w = 0.6$ two trajectories of pair approximation dynamics given by eqns (B.4) in Appendix B are illustrated in Fig. 5(b). Figure 5(c) shows the change in ρ_T for different w . This indicates that AD is evolutionarily stable when w is less than 0.77 and that TFT is evolutionarily stable when w is larger than 0.49.

The prediction of pair approximation dynamics for a two-dimensional model in Fig. 5(c) is intermediate between the complete mixing case (Fig. 1) and the one-dimensional case [Fig. 3(c)]. This can be understandable as the number of neighbors is two for the one-dimensional lattice model, eight for the two-dimensional lattice, and infinitely large for the completely mixing model.

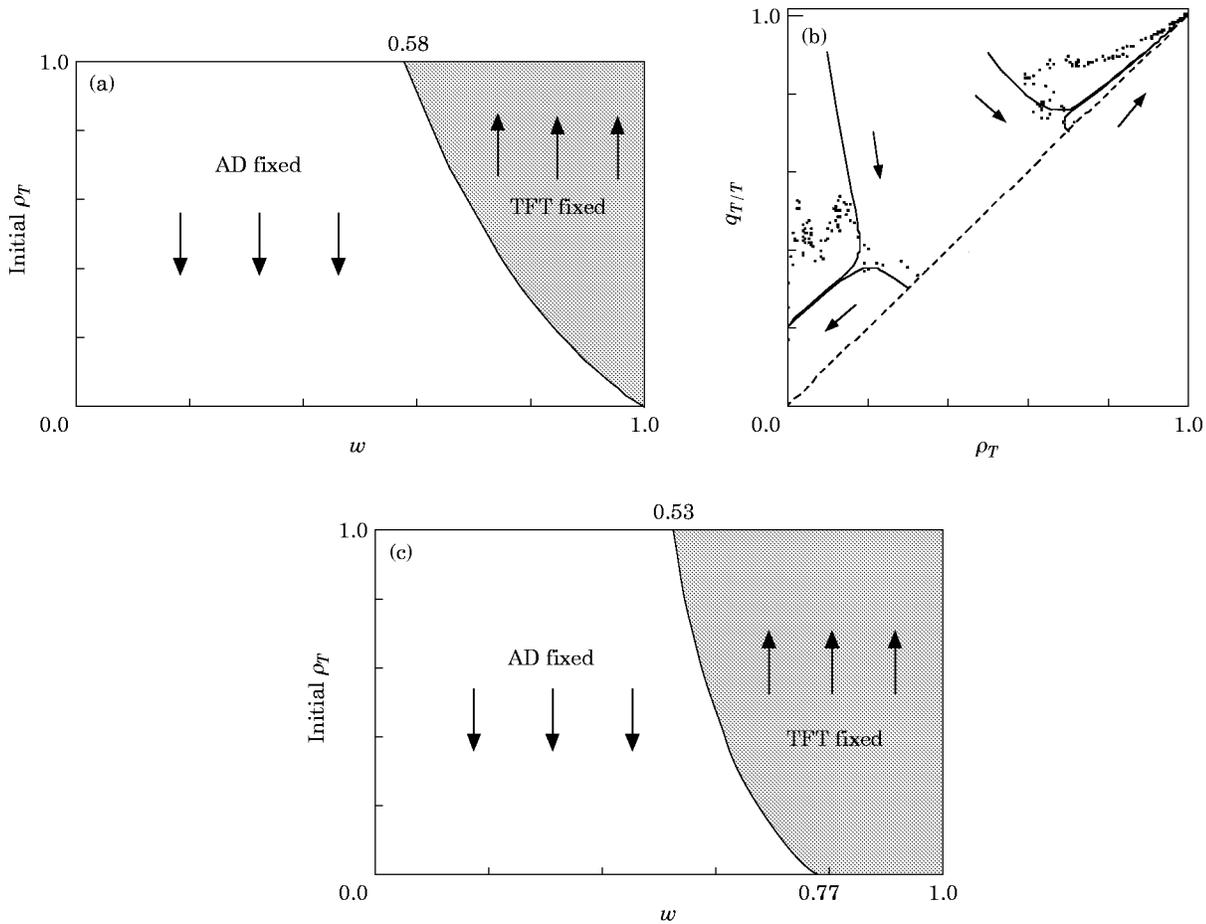


FIG. 5. Analyses of the two-dimensional lattice model. (a) The phase plane of the dynamics based on mean-field approximation, eqn (B.2). Horizontal axis is the probability of reiteration w , vertical axis is the density of TFT ρ_T . The region in which ρ_T increases is indicated by shade, given by eqn (B.2). This predicts that AD is always evolutionarily stable. The system shows bistability when w is larger than w_d . (b) Curves are the trajectories of the dynamics of pair approximation when w is 0.6. Dots are from computer simulation where the time interval between points is 10. The horizontal axis is global density ρ_T and the vertical axis is local density q_T/τ . A broken line is for $q_T/\tau = \rho_T$ implying a random spatial pattern. Arrows indicate the direction of movement along trajectories. The system is bistable. (c) The phase plane of the dynamics based on pair approximation, eqns (B.4). Horizontal axis is the probability of reiteration w , vertical axis is the density of TFT ρ_T . The region in which ρ_T increases is indicated by shade, given by eqns (B.4). When w is less than 0.77, AD is an evolutionarily stable strategy. When w is larger than about 0.49, TFT is an evolutionarily stable strategy. When the range of w is between 0.49 and 0.77, whether TFT or AD will be fixed in the lattice depends on ρ_T . The prediction is intermediate between the complete mixing case (Fig. 1) and one-dimensional case [Fig. 3(c)]. Note that the rare TFT can invade the population composed of AD if probability of iteration w is sufficiently high.

COMPUTER SIMULATION

The method of computer simulation of a two-dimensional square lattice model ($z = 8$, with Moore neighborhood) was the same as for the one-dimensional model explained before. We assumed a periodic boundary condition: i.e. the lattice was a torus (the rightmost column is the nearest neighbor of the leftmost column, and the top row is the nearest neighbor of the bottom row). We have carried out simulations both on a lattice of size 20×20 and on a lattice of size 100×100 . We normally computed the model until a time at which one of the two types is fixed in the population. If the system is bistable, the

outcome depends on the initial density. The initial population is of a random spatial pattern with different fractions of TFT: we computed ρ_T with 0.1, 0.3, 0.5, 0.7 and 0.9.

Examples of spatial patterns of the model on the lattice of 20×20 are illustrated in Fig. 6(a–c). The phase plane on $w - \rho_T$ for the lattice of size 20×20 is given in Fig. 6(d). TFT is the evolutionarily stable strategy for w larger than a critical value which lies between 0.5 and 0.6. AD is the evolutionarily stable strategy for w less than another threshold between 0.6 and 0.8. For w between these two critical values, this system shows bistability and the evolutionary outcome depends on the initial global density of TFT.

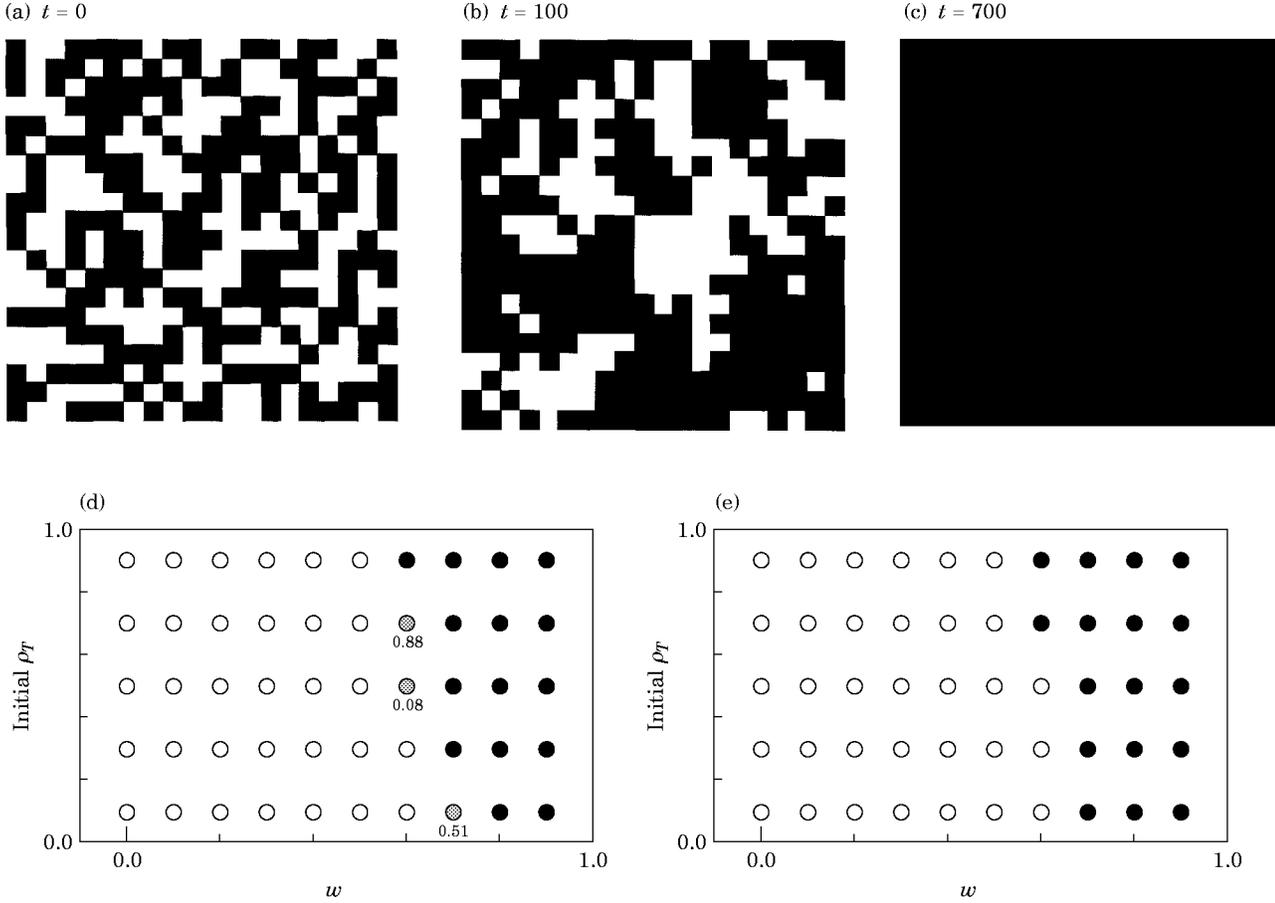


FIG. 6. (a–c) The spatial patterns generated by the two-dimensional lattice model with $w = 0.8$ and initial density $\rho_T = 0.5$ when lattice size is 20×20 . (a) Initial random distribution ($t = 0$), (b) patchy distribution is quickly formed ($t = 100$), and (c) TFT is fixed in the end ($t = 700$). (d) This illustrates change in initial ρ_T with time for different w when lattice size is 20×20 . Open circles, shaded circles and solid circles in the graph show the computer simulations of AD fixed, either AD or TFT fixed and TFT fixed in the lattice, respectively. The numbers below shaded circles are the fractions of runs in which TFT become fixed. When w is less than about 0.8, AD is an evolutionarily stable strategy. When w is larger than about 0.4, TFT is an evolutionarily stable strategy. For w between 0.5 and 0.8, the system shows bistability, i.e. whether or not TFT is fixed depends on the initial global density of TFT. The computer simulation results are explained well by the dynamics based on pair approximation [Fig. 5(c)]. (e) This illustrates change in initial ρ_T with time for different w when lattice size is 100×100 .

For several sets of parameters indicated by grey circles, some runs end up with fixation of TFT but other runs end up with fixation of AD. This is caused by stochasticity due to the finiteness of the lattice size.

The phase plane on $w - \rho_T$ for a large lattice of 100×100 is in Fig. 6(e). The parameter region showing fixation of TFT and fixation of AD are quite similar, but grey circles are absent. Depending on the parameter w and initial condition ρ_T , the evolutionary outcome is either fixation of TFT (solid circles) or fixation of AD (open circles). Bistability was observed for the case of $w = 0.6$ —TFT is fixed in all 50 runs if the initial ρ_T is 0.7 and 0.9, but AD is fixed if the initial ρ_T is 0.5 or smaller.

The results of the computer simulation, in Fig. 6(d) and 6(e), are again consistent with the pair approximation dynamics [as shown in Fig. 5(c)], but

not with the mean-field approximation [Fig. 5(a)]. The mean-field dynamics predict that AD always repels the invasion by rare TFT. However, according to the computer simulation, rare TFT can invade an AD population if w is larger than 0.7, which is similar to the 0.77 predicted by the pair-approximation dynamics.

We must note, however, our results of computer simulation on lattices of finite size (20×20 and 100×100) are different from behavior predicted for an infinitely large lattice, as there is a general mathematical argument that no bistability is possible on an infinitely large lattice irrespective of the dimensionality (Liggett, 1978; Durrett, 1980; Gray, 1982). The arguments can be explained intuitively as follows: For bistability to occur in the system, we must consider the situation where one type is

dominating all the local areas in the system. However, because the model is stochastic, and because the lattice size is infinitely large, there will always appear an area in which one of the two types dominates, and another area in which the other type dominates. Then, whether one type or the other wins, is determined simply by the movement of the boundary between two areas, in each of which one of the two types dominates.

Hence, if we analyse an infinitely large lattice, the phase plane should show no bistability, and the prediction by pair approximation, such as in Fig. 5(c), is not valid. It will look more like Fig. 3(c). Pair approximation assumes that the correlation between sites can be approximated considering correlation between neighbors. It is quite plausible that pair approximation fails if the whole system is segregated into large subareas in which the relative density of TFT and AD greatly differ and long range spatial correlation of states is large.

The computer simulation of the model on a finite size lattice showed bistability, which should disappear if we examine an infinitely large lattice.

On the other hand, if in biological, social situations there are a finite number of lattice sites, then the model on a lattice of reasonably large but still finite size, may be more suitable as a modeling tool. The behavior of this lattice was closely described by pair approximation dynamics.

Discussion

To evaluate the effect of spatial structure of populations, we need to compare a lattice model and models with randomized spatial configuration in which the ecological interactions and fitness functions are the same. From mortality, eqn (4), the fitness corresponding to lattice models is given by an exponential function of the total score as:

$$(\text{fitness}) = 1 - K \exp[-\alpha(\text{total score})] \quad (14)$$

where K and α are positive constants. The total score is given by the interaction with z neighbors. The evolutionary game using the fitness eqn (14) with a spatially randomized configuration is comparable to the other models in this paper, and we call this as “the complete mixing” model in the following. In Appendix B, we show that the complete mixing model with eqn (14) is the same as a simpler model with a linear fitness given by eqn (1) in the limit when α is very small. Hence we can regard Fig. 1 as the phase plane for the complete mixing model when α is small. Importantly, the phase plane is independent of the number of neighbors with which each player interacts,

z , or of the dimensionality of the lattice (see Appendix B for derivation).

The mean-field approximation [Fig. 3(a)] for the one-dimensional model fails to explain the result of computer simulation, especially the absence of bistability [Fig. 2(b)], but pair approximation dynamics [Fig. 3(c)] are quite accurate, and are also consistent with the invadability analysis. The same conclusion holds for the two-dimensional model [Fig. 5(c) is close to Fig. 6(d)]. This again confirms the usefulness of pair-approximation for lattice structured models (Sato *et al.*, 1992; Harada & Iwasa, 1994; Harada *et al.*, 1995).

The problem of the difficulty in invasion and establishment of the cooperative strategy in a population dominated by a noncooperative one can be resolved in the lattice structured population. Both in the one-dimensional and in the two-dimensional lattices, initially rare TFT can increase in an AD-dominated population and replace AD if the iteration probability w is sufficiently high [Fig. 3(c), Fig. 5(c)]. In the lattice population, cooperation can evolve from an initially noncooperative society.

However, TFT is evolutionarily stable only when w is larger than 0.6 in the one-dimensional lattice model, while it is so when w is larger than 0.5 in the complete mixing model. Comparing the shaded area of the $w - \rho_T$ phase plane for the increase of TFT given in Fig. 1 for the complete mixing model and the corresponding area in Fig. 3(c) for the one-dimensional lattice model, we see that neither one contains the other inside. Lattice structure facilitates the evolution of cooperation when TFT, cooperative strategy, is rare, but inhibits it when TFT is common.

The condition for the evolution of cooperation predicted by mean-field dynamics is much more stringent than the one observed in the computer simulation or given by pair approximation. Especially notable is that in a one-dimensional lattice, the shaded area of the phase plane leading to the ultimate fixation of TFT in Fig. 3(a) is much smaller than the one by the complete mixing calculation (Fig. 1). This is puzzling because both the mean-field approximation and the complete mixing calculation assume perfectly randomized spatial configuration. This difference can be explained by considering two population processes in which mean-field dynamics and the complete mixing model, respectively, are exactly valid. First we consider a population in which the location of TFT and AD are randomized. Players interact with their neighbors, their mortality depends on the score. After the death of a player before filling the vacant site, we may recalculate the fraction of TFT and AD and generate a new randomly

distributed spatial pattern using this new fraction. This stochastic process exactly follows the complete mixing model.

We may imagine the second process in which each vacant site produced by the death of a player is immediately filled by a copy of randomly chosen neighbors, then we randomize the spatial configuration. This corresponds to the mean-field dynamics. In this second process, any behavior to reduce the score of its neighbors enhances the mortality of neighbors, and can be advantageous to the player who enjoys a higher opportunity to reproduce the vacant sites produced by the death of the neighbor. The major difference between the complete mixing model and the mean-field dynamics lies in the possible advantage of such “spiteful” behavior, which explains why the spread of TFT is much more difficult under mean-field dynamics [Fig. 3(a)] than in the complete mixing dynamics (Fig. 1).

Now we compare the mean-field dynamics [Fig. 3(a) for one-dimensional model, and Fig. 5(a) for the two-dimensional model] and the lattice models [Fig. 3(c) and Fig. 5(c), respectively]. Both allow some advantage to spiteful behaviors. However, neighbors are independently chosen in mean-field dynamics, but are similar in lattice models. This difference in the nearest-neighbor correlation explains why the shaded area of the phase plane for the latter is larger than that for the former.

The lattice models [Fig. 3(c) or Fig. 5(c)] are different from the complete mixing model (Fig. 1) on two points: the possible advantage of being spiteful makes cooperation less likely to evolve and the correlation between neighbors makes it more likely to evolve. The possibility of spiteful behavior in the lattice model discouraging the evolution of altruism has been pointed out previously mostly from computer simulation studies (Matsuda, 1987; Matsuda *et al.*, 1987; Taylor, 1992; Wilson *et al.*, 1992). In a lattice population the movement of the boundary between a TFT cluster and an AD cluster determines the fate of evolution. The total fraction of TFT in the whole population does not affect the relative advantage of two strategies.

A biological example of spiteful behavior is toxin production by bacteria. Some strains of the bacteria *E. coli* produce a toxin, called colicin, but they themselves are immune to it. However, colicin-producing strains have a slower rate of population growth than colicin-sensitive strains. Durrett & Levin (manuscript in review) studied the coexistence of colicin-producing and colicin-sensitive strains in a lattice, as they cannot coexist in a perfectly mixed population. The model is similar to the nearest-neighbor-

migration model for the evolution of altruism by Matsuda (1987).

Matsuda (1987) and Matsuda *et al.* (1987) studied the extinction-invasion lattice model in which the dynamics of three states (sites occupied by the cooperative type, those occupied by selfish type and vacant sites) are examined. In the absence of kin-recognition, cooperation with neighbors can evolve if the migration range is small and the habitat is not saturated (i.e. many sites are vacant). However, this effect was not very strong if most sites are occupied. The reason is that a high score for cooperative individuals in the middle of tight clusters would not contribute to their spread. Similarly, in our model TFT forms tight clusters and vacant sites within a cluster of AD would be filled by AD even if they have low scores. What matters is the movement of the boundary between a tight cluster of TFT and a neighboring cluster of AD. However, the model studied in the current paper is simpler than the one studied by Matsuda, because we assumed only two states (sites occupied by TFT and those occupied by AD) neglecting the possibility that the site may be vacant.

Wilson *et al.* (1992) studied a lattice model with similar to our two states, but the mode of reproduction was different. Wilson *et al.* assumed that the death of an individual occurs randomly and that the vacant site is subsequently filled by a copy of neighbors chosen with scores (or fitnesses) considered. Wilson *et al.* (1992) and Taylor (1992) indicates that the local population regulation limits the evolution of altruism in purely viscous populations.

In the present paper, we assumed that the survivorship of a player depends on its score but the reproductive performance is independent of the score. In contrast, in other models (e.g. Axelrod, 1984; Nowak & May, 1992; Wilson *et al.*, 1992), the score affects the likelihood of being copied (i.e. reproductive rate), rather than the longevity. A biological example of score-dependent mortality may be “allelopathy”, the production of chemicals by plants and bacteria that harm their neighboring competitors, while an example of score-dependent fertility may be the enhancement of seed production of neighboring plants by attracting pollinators in self-incompatible flowers. In general, many modes of social interaction, such as competition for light among plants, would affect both mortality and fertility.

This difference in the assumption may cause qualitative differences in the model’s behavior. For example, Nowak & May (1992) observed the

perpetual coexistence of cooperative and defective players, forming constantly changing spatial patterns for a particular set of parameters. Coexistence of the two different strategies is not observed in models studied in this paper. It would be an interesting theme of future study to examine whether this difference comes from the assumption of score-dependent fertility in Nowak & May and score-dependent mortality in the current model, or from the other details such as synchronous vs asynchronous state changes. Although Wilson *et al.* (1992) studied score-dependent and random mortality, whether or not the coexistence of two types was possible is not very clear from their figures explaining computer simulations.

Games played on the lattice may be usefully extended. For example, an N-person Prisoner's Dilemma game has been intensively studied in social psychology (Hamburger, 1973; Dawes, 1980), in which more than two players interact. An N-person game played in a lattice structured population is of considerable interest.

Second, a recent study of the iterated Prisoner's Dilemma game reveals that if players make "errors" with a small probability, then strategies other than TFT can be better, such as Pavlov (the strategy that cooperates if and only if both players opted for the same alternative in the previous move) (Nowak & Sigmund, 1993) and GTFT (the strategy that always cooperates when its partner cooperated in the previous move and cooperates with probability 0.9 when its partner defected in the previous move) (Nowak & Sigmund, 1992). These simulation results were supported by analytical study (Stephens *et al.*, 1995). These studies on the effect of errors have been carried out in the complete mixing populations. It would also be interesting to examine the evolutionary outcome of these diverse strategies in a lattice-structured population.

This work is supported in part by a Grant-in-Aid Scientific Research by Ministry of Education, Science and Culture, Japan. We thank Professor Ken-iti Sato, Graduate School of Human Informatics, Nagoya University, for his advice and support. We also thank M. Boots, R. Durrett, Y. Haraguchi, T. Kubo, T. Kuroda, S. A. Levin, H. Morimoto, T. Yamagishi and K. Yamaguchi for their helpful comments.

REFERENCES

- ALEXANDER, R. D. (1979). *Darwinism and Human Affairs*. Seattle: University of Washington Press.
- AXELROD, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- AXELROD, R. & HAMILTON, W. D. (1981). The evolution of cooperation. *Science* **211**, 1390–1396.
- BOYD, R. (1989). Mistakes allow evolutionary stability in the repeated prisoner's dilemma game. *J. theor. Biol.* **136**, 47–56.
- BOYD, R. & LORBERBAUM, J. P. (1987). No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game. *Nature* **327**, 58–59.
- CASWELL, H. & ETTER, R. J. (1992). Ecological interactions in patchy environments: from patch-occupancy models to cellular automata. In: *Patch Dynamics: Lecture Notes in Biomathematics* (Levin, S. A., Powell, T. M. & Steele, J. H., eds) Vol. 96, pp. 93–109. New York: Springer-Verlag.
- DAWES, R. M. (1980). Social Dilemmas. *Ann. Rev. Psychol.* **31**, 169–193.
- DE WAAL, F. B. M. (1982). *Chimpanzee Politics: Power and Sex Among Apes*. New York: Harper and Row.
- DURRETT, R. (1980). On the growth of one dimensional contact processes. *Ann. Prob.* **8**, 890–907.
- DURRETT, R. & LEVIN, S. A. (1994a). Importance of being discrete (and spatial). *Theoret. Popul. Biol.* **46**, 363–394.
- DURRETT, R. & LEVIN, S. A. (1994b). Stochastic spatial models: a user's guide to ecological applications. *Phil. Trans. R. Soc. Lond. B* **343**, 329–350.
- GRAY, L. F. (1982). The positive rates problem for attractive nearest neighbor spin system on z . *Zeit. für Wahrscheinlichkeitstheorie* **61**, 389–404.
- HAMBURGER, H. (1973). N-person prisoner's dilemma. *J. Math. Sociol.* **3**, 27–48.
- HAMILTON, W. D. (1964). The genetical evolution of social behavior. *J. theor. Biol.* **7**, 1–52.
- HARADA, Y. & IWASA, Y. (1994). Lattice population dynamics for plants with dispersing seeds and vegetative propagation. *Res. Popul. Ecol.* **36**, 237–249.
- HARADA, Y., EZOE, H., IWASA, Y., MATSUDA, H. & SATO, K. (1995). Population persistence and spatially limited social interaction. *Theoret. Popul. Biol.* **48**, 65–91.
- LIGGETT, T. M. (1978). Attractive nearest neighbor spin systems on the integers. *Ann. Prob.* **6**, 629–636.
- NOWAK, M. A. & MAY, R. M. (1992). Evolutionary games and spatial chaos. *Nature* **359**, 826–829.
- NOWAK, M. A. & SIGMUND, K. (1992). Tit for tat in heterogeneous populations. *Nature* **355**, 250–253.
- NOWAK, M. A. & SIGMUND, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* **364**, 56–58.
- MATSUDA, H. (1987). Condition for the evolution of altruism. In: *Animal Societies: Theories and Facts* (Ito, Y., Brown, J. L. & Kikkawa, J., eds) pp. 67–80. Tokyo: Japan Sci. Soc. Press.
- MATSUDA, H., TAMACHI, N., OGITA, N. & SASAKI, A. (1987). A lattice model for population biology. In: *Mathematical Topics in Biology: Lecture Notes in Biomathematics* (Teramoto, E. & Yamaguti, M., eds) Vol. 71, pp. 154–161. New York: Springer-Verlag.
- MATSUDA, H., OGITA, N., SASAKI, A. & SATO, K. (1992). Statistical mechanics of population—the lattice Lotka–Volterra model. *Prog. Theor. Phys.* **88**, 1035–1049.
- SATO, K., MATSUDA, H. & SASAKI, A. (1994). Pathogen invasion and host extinction in lattice structured populations. *J. Math. Biol.* **32**, 251–268.
- STEPHENS, D. W., NISHIMURA, K. & TOYER, K. B. (1995). Error and discounting in the iterated prisoner's dilemma. *J. theor. Biol.* **176**, 457–469.
- TAYLOR, P. D. (1992). Altruism in viscous populations—an inclusive fitness model. *Evol. Ecol.* **6**, 352–356.
- TRIVERS, R. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57.
- WILKINSON, G. S. (1984). Reciprocal food sharing in the vampire bat. *Nature* **308**, 181–184.
- WILSON, D. S., POLLOCK, G. B. & DUGATKIN, L. A. (1992). Can altruism evolve in purely viscous populations? *Evol. Ecol.* **6**, 331–341.

APPENDIX A

The probability that a cluster consisting of n TFT goes extinct ultimately is denoted by Q_n . Let $\lambda_n \Delta t$ be the transition probability in a short time interval of Δt from a cluster consisting of n TFT to a cluster with $(n+1)$ TFT. Let $\mu_n \Delta t$ be the transition probability from n TFT to $(n-1)$ TFT. Now the basic equation of Random-Walk is:

$$Q_n = Q_n(1 - \lambda_n \Delta t - \mu_n \Delta t) + Q_{n+1} \lambda_n \Delta t + Q_{n-1} \mu_n \Delta t. \quad (\text{A.1})$$

We first consider the case in which the cluster will not go extinct once the cluster reaches a sufficiently large size $k+1$. Later we make k infinitely large. Then the boundary conditions are:

$$Q_n = 0 \quad (n \geq k-1) \quad (\text{A.2a})$$

$$Q_0 = 1. \quad (\text{A.2b})$$

From the definition of the processes, the transition occurs when a site at the boundary between a run of T-sites and a run of D-sites dies and subsequently replaced by its neighbor. The transition rates, λ_n and μ_n , can be rewritten as follows:

$$\lambda_n = \begin{cases} 0 & n = 0 \\ M_{D,1} & n = 1, 2, 3, \dots \end{cases} \quad (\text{A.3a})$$

where $n=0$ is an absorbing point

$$\mu_n = \begin{cases} M_{T,0} & n = 1 \\ M_{T,1} & n = 2, 3, 4, \dots \end{cases}. \quad (\text{A.3b})$$

In the following we put $\lambda = M_{D,1}$, $\mu = M_{T,1}$ and $\mu' = M_{T,0}$.

Using eqns (A.2a–A.3b), eqn (A.1) can be rewritten as:

$$Q_n = Q_n(1 - \lambda \Delta t - \mu \Delta t) + Q_{n+1} \lambda \Delta t + Q_{n-1} \mu \Delta t \quad 2 \leq n \leq k \quad (\text{A.4a})$$

$$Q_1 = Q_1(1 - \lambda \Delta t - \mu' \Delta t) + Q_2 \lambda \Delta t + Q_0 \mu \Delta t \quad n = 1. \quad (\text{A.4b})$$

From eqns (A.2a) and (A.4a), we have

$$Q_{n+1} = Q_1 + \sum_{k=1}^n (\mu/\lambda)^{k-1} (Q_2 - Q_1). \quad (\text{A.5})$$

From eqns (A.2b) and (A.4b), we have

$$Q_2 = -\frac{\mu'}{\lambda} + \frac{\lambda + \mu'}{\lambda} Q_1. \quad (\text{A.6})$$

From eqns (A.2a), (A.5) and (A.6), Q_1 is derived as:

$$Q_1 = \frac{\mu'(1 - (\mu/\lambda)^k)}{\lambda - \mu + \mu'(1 - (\mu/\lambda)^k)} \quad \text{if } \lambda \neq \mu \quad (\text{A.7a})$$

$$Q_1 = \frac{\mu'k}{\lambda + \mu'k} \quad \text{if } \lambda = \mu. \quad (\text{A.7b})$$

$\lambda = \mu$ occurs when $w = w_b = (T - S - R + P)/(T - S)$.

From eqns (A.7a) and (A.7b), when $k \rightarrow \infty$, Q_1 converges to the following values:

$$Q_1 \xrightarrow{k \rightarrow \infty} \frac{\mu'}{\mu' + \lambda - \mu} = \frac{1}{1 + e^{-\alpha(P+T-2S)} - e^{-\alpha\left(P-S+\frac{R-P}{1-w}\right)}} \quad \text{if } \mu < \lambda (w > w_b), \quad (\text{A.8a})$$

$$Q_1 \xrightarrow{k \rightarrow \infty} 1 \quad \text{if } \mu \geq \lambda (w \leq w_b), \quad (\text{A.8b})$$

where we rewrote λ , μ and μ' in terms of pay-offs. Equation (A.8) indicates that if the rate of increase in n is larger than the rate of decrease $\lambda > \mu$, then there is a positive probability for TFT not to go extinct ultimately. The probability TFT survive in the lattice starting from a single TFT is $Q_T = 1 - Q_1$, which is in Fig. 4.

In a similar way, the probability, R_n , that a colony consisting of n AD goes extinct is also derived as follows:

$$\lambda_n = \begin{cases} 0 & n = 0 \\ M_{T,1} & n = 1, 2, 3, \dots \end{cases}, \quad (\text{A.9a})$$

$$\mu_n = \begin{cases} M_{D,2} & n = 1 \\ M_{D,1} & n = 2, 3, 4, \dots \end{cases}, \quad (\text{A.9b})$$

where $\lambda = M_{T,1}$, $\mu' = M_{D,2}$ and $\mu = M_{D,1}$. We have recurrence formulae as:

$$R_n = R_n(1 - \lambda \Delta t - \mu \Delta t) + R_{n+1} \lambda \Delta t + R_{n-1} \mu \Delta t \quad 2 \leq n \leq k \quad (\text{A.10a})$$

$$R_1 = R_1(1 - \lambda \Delta t - \mu' \Delta t) + R_2 \lambda \Delta t + R_0 \mu \Delta t \quad n = 1. \quad (\text{A.10b})$$

Then

$$R_1 \xrightarrow{k \rightarrow \infty} \frac{1}{1 - e^{\alpha(T-P)} + e^{\alpha\left((2R-P-S)-\frac{R-P}{1-w}\right)}} \quad \text{if } \mu < \lambda (w > w_b), \quad (\text{A.11a})$$

$$R_1 \xrightarrow{k \rightarrow \infty} 1 \quad \text{if } \mu \geq \lambda (w \leq w_b). \quad (\text{A.11b})$$

The probability for a newly invading AD to survive in the lattice is $R_D = 1 - R_1$, which is shown in Fig. 4.

APPENDIX B

(i) The Model on a Two-dimensional Lattice

Equation (13) is rewritten as:

$$\begin{aligned} \frac{d\rho_T}{dt} = & -\rho_T \sum_{n=0}^{z-1} \binom{z}{n} (q_{T|T})^n (q_{D|T})^{z-n} M_{T,n} \frac{(z-n)}{z} \\ & + \rho_D \sum_{n=1}^z \binom{z}{n} (q_{T|D})^n (q_{D|D})^{z-n} M_{D,n} \frac{n}{z}. \end{aligned} \quad (\text{B.1})$$

If we adopt mean-field approximation, eqn (B.1) is rewritten as:

$$\begin{aligned} \frac{d\rho_T}{dt} = & \frac{1}{8} \rho_T (1 - \rho_T) \\ & \times \left\{ -\sum_{n=0}^7 \binom{z}{n} \rho_T^n (1 - \rho_T)^{7-n} M_{T,n} (8-n) \right. \\ & + (1 - \rho_T)^{-7} \sum_{n=1}^8 \binom{z}{n} (\rho_T (1 - \rho_T))^{n-1} \\ & \left. \times (1 - 2\rho_T + \rho_T^2)^{8-n} M_{D,n} n \right\}. \end{aligned} \quad (\text{B.2})$$

To derive the equations for local density based on pair approximation, we first derive the dynamics of doublet density as:

$$\begin{aligned} \frac{d\rho_{TT}}{dt} = & \frac{2}{z} \left\{ -\rho_T \sum_{n=1}^{z-1} n \binom{z}{n} (q_{T|T})^n (q_{D|T})^{z-n} M_{T,n} \frac{(z-n)}{z} \right. \\ & \left. + \rho_D \sum_{n=1}^z n \binom{z}{n} (q_{T|D})^n (q_{D|D})^{z-n} M_{D,n} \frac{n}{z} \right\} \end{aligned} \quad (\text{B.3})$$

where the first term indicates a transition of TFT indicating with n TFT and $(z-n)$ AD to AD, with the probability $(z-n)/z$ (this transition changes n pairs of TT to $(n-1)$ pairs), and the second term indicates a transition of AD interacting with n TFT and $(z-n)$ AD to TFT, with the probability n/z (this transition changes n pairs of TT to $(n+1)$ pairs).

By adopting pair-approximation, we can construct a closed dynamical system of ρ_T and $q_{T|T}$ with $z = 8$ as:

$$\begin{aligned} \frac{d\rho_T}{dt} = & \frac{1}{8} \rho_T (1 - q_{T|T}) \\ & \times \left\{ -\sum_{n=0}^7 \binom{8}{n} q_{T|T}^n (1 - q_{T|T})^{7-n} M_{T,n} (8-n) \right. \\ & + (1 - \rho_T)^{-7} \sum_{n=1}^8 \binom{8}{n} (\rho_T (1 - q_{T|T}))^{n-1} \\ & \left. \times (1 - 2\rho_T + \rho_T q_{T|T})^{8-n} M_{D,n} n \right\} \end{aligned} \quad (\text{B.4a})$$

$$\begin{aligned} \frac{dq_{T|T}}{dt} = & \frac{1}{8} (1 - q_{T|T}) \\ & \times \left\{ -\frac{1}{4} \sum_{n=0}^7 \binom{8}{n} q_{T|T}^n (1 - q_{T|T})^{7-n} M_{T,n} n (8-n) \right. \\ & + \frac{1}{4} (1 - \rho_T)^{-7} \sum_{n=1}^8 \binom{8}{n} (\rho_T (1 - q_{T|T}))^{n-1} \\ & \times (1 - 2\rho_T + \rho_T q_{T|T})^{8-n} M_{D,n} n^2 \\ & + q_{T|T} \sum_{n=0}^7 \binom{8}{n} q_{T|T}^n (1 - q_{T|T})^{7-n} M_{T,n} (8-n) \\ & \left. - q_{T|T} (1 - \rho_T)^{-7} \sum_{n=1}^8 \binom{8}{n} (\rho_T (1 - q_{T|T}))^{n-1} \right. \\ & \left. \times (1 - 2\rho_T + \rho_T q_{T|T})^{8-n} M_{D,n} n \right\}. \end{aligned} \quad (\text{B.4b})$$

(ii) The Complete Mixing Game with Fitness Given by eqn (14)

The fitnesses for TFT and AD are:

$$W_T = 1 - K \exp[-\alpha B_{T,n}] \quad (\text{B.5a})$$

$$W_D = 1 - K \exp[-\alpha B_{D,n}] \quad (\text{B.5b})$$

respectively. Then using eqn (1) and Table 2 for scores, we have

$$\begin{aligned}
 W_T &= \sum_{n=0}^z \binom{z}{n} (1 - K \exp[-\alpha(nV(T/T) \\
 &\quad + (z-n)V(T/D))]) \rho_T^n \rho_D^{z-n} \\
 &= 1 - K(e^{-\alpha V(T/T)} \rho_T + e^{-\alpha V(T/D)} \rho_D)^z. \quad (\text{B.6a})
 \end{aligned}$$

Similarly, we have

$$W_D = 1 - K(e^{-\alpha V(D/T)} \rho_T + e^{-\alpha V(D/D)} \rho_D)^z. \quad (\text{B.6b})$$

The system can be bistable. A curve on $w - \rho_T$ plane separating two domains of attraction is given by $W_T = W_D$, which is rewritten as

$$\begin{aligned}
 \rho_T &= \frac{e^{-\alpha V(D/D)} - e^{-\alpha V(T/D)}}{e^{-\alpha V(T/T)} - e^{-\alpha V(T/D)} + e^{-\alpha V(D/D)} - e^{-\alpha V(T/D)}} \\
 &= \frac{e^{-\alpha S} - e^{-\alpha P}}{e^{-\alpha S} - e^{-\alpha P} + e^{-\alpha T} - e^{-\alpha(T-wp)/(1-w)}}. \quad (\text{B.7})
 \end{aligned}$$

Note that the curve is independent of z , the number of neighbors for each site, and hence the dimensionality of the lattice. In the limit when α is very small, eqn (B.7) becomes eqn (2), for the complete mixing model.