Assortative clustering in a one-dimensional population with replication strategies

Sunhee Chae,¹ Nahyeon Lee,¹ Seung Ki Baek,²† and Hyeong-Chai Jeong†¹

¹Department of Physics and Astronomy, Sejong University, Seoul 05006, Korea
²Department of Physics, Pukyong National University, Busan 48513, Korea

(Dated: March 16, 2021)

In a geographically distributed population, assortative clustering plays an important role in evolution by modifying local environments. To examine its effects in a linear habitat, we consider a one-dimensional grid of cells, where each cell is either empty or occupied by an organism whose replication strategy is genetically inherited to offspring. The strategy determines whether to have offspring in surrounding cells, as a function of the neighborhood configuration. If more than one offspring compete for a cell, then they can be all exterminated due to the cost of conflict depending on environmental conditions. We find that the system is more densely populated in an unfavorable environment than in a favorable one because only the latter has to pay the cost of conflict. This observation agrees reasonably well with a mean-field analysis which takes assortative clustering of strategies into consideration. Our finding suggests a possibility of intrinsic nonlinearity between environmental conditions and population density when an evolutionary process is involved.

I. INTRODUCTION

“Space exists so that everything doesn’t happen to you,” says Susan Sontag. Spatiality often means being exempt from interacting with all others: One may be surrounded by more favorable neighbors than the average or the opposite, when the spatial configuration is nonuniform. If the local environments experienced by individuals differ from place to place, then it implies different selection pressure in terms of evolution, which can shape the local environments even more differently. If such a feedback loop forms, then individuals can break away from the evolutionary path that would have been followed in a well-mixed population. For this reason, the roles of spatiality in evolution have been studied extensively in the literature [1–5].

To be more specific, let us consider a model of cellular automata, one of the simplest models of life in spatial dimensions, yet with the possibility of genuine complexity in its behavior [6–12]. In a cellular-automata model, the space is divided into discrete cells, and the cells can be occupied by “organisms” that replicate themselves according to mechanistic laws. To put such a model into an evolutionary context, we would like to point out the following: The replication process would generate different copies with small errors in practice, and each of the different copies would also have different efficiency in replicating itself. In other words, they must be subject to an evolutionary process of mutation and selection.

In this work, we will study evolution of such cellular organisms in silico by assigning a replication strategy to each of them. The strategy is transmitted genetically to offspring, and it has to compete with others in neighboring cells. This defines a game in the sense of game theory because an organism’s payoff, identified with the number of offspring, will depend on its neighbors’ strategies as well as on its own. An aggressive strategy would produce as many offspring as possible, invading the territories of other strategies. Even if it incurs extra cost of conflict and thus reduces the total size of the population, it should have a higher chance to spread than nonaggressive ones. If we regard the total population growth as the collective interest of life, then it thus conflicts, at least partially, with individual interests of the selfish genes that encode replication strategies. However, a paradox of evolution is that self-interested behavior is not always favored by selection [13–14], provided that the dynamical rule permits assortative clustering of players who conform to collective interests [15–19]. This study will show that such an assortative effect can be induced in a spatial game by a simple mechanism, whereby deflection from collective interests is successfully suppressed. As a consequence, the mechanism introduces nonlinearity in the relation between environmental conditions and population density.

This work is organized as follows: In the next section, we introduce our model. The Monte Carlo simulation result will be presented in Sec. III. After explaining the observed behavior with a mean-field approximation in Sec. IV, we conclude this work in Sec. V.

II. MODEL

Let us consider a group of organisms living on a one-dimensional grid with the periodic-boundary conditions to see the assortative effect most clearly. In ecology, such a one-dimensional structure describes a habitat constrained by linear environmental features such as rivers or shorelines [20, 21], and it is also physically relevant to studying dynamic processes in (1+1) dimensions [22, 23]. Each grid cell is indexed by \(x\), and its occupancy is denoted by \(n_x\): It can be either empty with \(n_x = 0\) or occupied by one of the organisms with \(n_x = 1\). Time \(t\) is also a discrete variable, under the assumption that the
organisms have nonoverlapping generations. The model consists of two parts, i.e., replication and mutation.

In the replication process, every organism produces an offspring with the same strategy in its own cell. At the same time, it may also produce offspring in neighboring cells. Therefore, at the beginning of a new generation, the number of offspring in a cell can sometimes be greater than 1. For example, let us imagine that only two neighboring cells, \( x - 1 \) and \( x \), are occupied in an otherwise empty system by organisms with strategies \( I \) and \( J \), respectively. However, as implied by \( n_x \leq 1 \), each cell can barely support a single adult: If the \( I \)-player at \( x - 1 \) produces two offspring \( i \) and \( i' \), one in its own cell and the other in a neighboring cell \( x \), then the latter will compete with the \( J \)-player's offspring \( j \) born in \( x \). By assumption, they all die with probability \( 1 - \alpha \), leaving the cell \( x \) empty, as a result of exhausting competition. Here we have introduced a parameter \( \alpha \) between 0 and 1, which can be interpreted as the favorability of the environment. With probability \( \alpha \), the cell remains occupied, i.e., \( n_x(t = 2) = 1 \), in which case we randomly choose one between \( i' \) and \( j \) as the survivor. If \( i' \) is chosen, then it will have grown into an \( I \)-player at \( t = 2 \); otherwise, we will have a \( J \)-player in \( x \) again. The above explanation can be schematically represented as follows:

\[
\begin{array}{c}
\text{t = 2} \\
I \quad \cdots \quad \Omega \\
i \quad i' \quad \rightarrow \quad j
\end{array}
\]

\[
\begin{array}{c}
\text{t = 1} \\
I \quad \cdots \quad J
\end{array}
\]

with

\[
\Omega = \begin{cases} 
E & \text{with prob. } 1 - \alpha^2 \\
I & \text{with prob. } \alpha^2/2 \\
J & \text{with prob. } \alpha^2/2 
\end{cases}
\]  

(1)

where \( E \) denotes that the cell is empty. Similarly, let us imagine that the system starts with only three organisms, which occupy three consecutive cells, \( x - 1, x, \) and \( x + 1 \), and play strategies \( I, J, \) and \( K \), respectively. If the \( I \)- and \( K \)-players produce their offspring in the \( J \)-player's cell \( x \), then the competition of the three will be more intense than the above case of two competitors. We describe this situation by assuming that the focal cell \( x \) becomes empty with probability \( 1 - \alpha^2 \), which is greater than \( 1 - \alpha \) for \( \alpha \in (0, 1) \). If the cell remains occupied, i.e., \( n_x(t = 2) = 1 \), one of the three competitors is chosen randomly as the survivor. This example can thus be represented as follows:

\[
\begin{array}{c}
\text{t = 2} \\
I \quad \cdots \quad \Omega \quad \cdots \quad K \\
i \quad i' \quad \rightarrow \quad j \quad \rightarrow \quad k
\end{array}
\]

\[
\begin{array}{c}
\text{t = 1} \\
I \quad \cdots \quad J \quad \rightarrow \quad K
\end{array}
\]

| neighboring-cell occupancy \( \nu \) | 11 10 01 00 |
| replication behavior \( \beta \) for each \( \nu \) | 00 01 10 11 |
| strategy index | 27 |

TABLE I: Example of a replication strategy indexed as 27, which produces offspring only in empty neighboring cells. Such behavior is characterized by \( \beta = \text{NOT } \nu \), where NOT means logical negation on each bit. As shown in the first line, we sort \( \nu \) in descending order from 11 to 00, so the binary representation of this strategy is obtained as 00011011 (the second row), which corresponds to 27 in decimal. With

\[
\Omega = \begin{cases} 
E & \text{with prob. } 1 - \alpha^2 \\
I & \text{with prob. } \alpha^2/2 \\
J & \text{with prob. } \alpha^2/2 \\
K & \text{with prob. } \alpha^2/2 
\end{cases}
\]  

(2)

When an organism exists in a cell \( x \), we assume that its replication strategy takes into account \( n_{x-1} \) and \( n_{x+1} \), that is, the occupancy of neighboring cells. We thus have to distinguish four cases, denoted by \( \nu_x = 2^1 \times n_{x-1} + 2^2 \times n_{x+1} \), so that \( \nu_x \) can take a value from \( \{0,1,2,3\} \). This variable can be conveniently represented in binary: If \( n_{x-1} = n_{x+1} = 1 \), for example, then we can write \( \nu_x = 11 \). Let \( b_{x \rightarrow y} \) be a binary variable for the replication behavior which represents whether the organism in \( x \) produces an offspring in \( y \): If it does, then \( b_{x \rightarrow y} = 1 \), and 0 otherwise. Note that \( b_{x \rightarrow x} = 1 \) because the organism will always produce an offspring in its own cell. Then the strategy of the organism in a cell \( x \) is determined by its replication behavior \( \beta_x = 2^1 \times b_{x \rightarrow x-1} + 2^2 \times b_{x \rightarrow x+1} \) as a function of \( \nu_x \in \{0,1,2,3\} \). The replication behavior \( \beta_x \) can also be represented in binary, for example, \( \beta_x(00) = 11 \) if \( b_{x \rightarrow x-1} = b_{x \rightarrow x+1} = 1 \) for \( \nu_x = 00 \). It means that the strategy will produce offspring in both the neighboring cells when they are empty. We now represent the strategy as an eight-digit binary number by arranging \( \beta_x(\nu_x) \) in descending order of \( \nu_x \) from 11 to 00. The most aggressive strategy will always produce offspring in the neighboring cells by assigning \( \beta_x = 11 \) to all four \( \nu_x \)'s. This strategy can thus be indexed as 11111111 in binary, which corresponds to 255 in decimal. Note that the subscript \( x \) can actually be dropped in the above description because the strategy itself has no dependence on the position. As another example, the most inactive strategy should have \( \beta = 00 \) for every \( \nu \in \{0,1,2,3\} \), hence 00000000 = 0 as its index, because it will never invade the neighboring cells. Among 256 possible strategies between these two extremes, Table I shows a nontrivial strategy that produces offspring only in empty neighboring cells, whose index is calculated as 27 in decimal. As will be shown by numerical simulation below, this turns out to be one of the most important strategies in our model.

In the presence of environmental noise, the strategic information may be lost in the course of replica-
Thus, we assume that an offspring’s strategy may change to an arbitrary one in the set of available strategies \( S = \{0, 1, \ldots, 255\} \) with small mutation probability \( \mu \ll 1 \). The mutation process is also important from a computational point of view: We will calculate time-averaged quantities from a Monte Carlo method. This would not be justified without mutation because the system might cease to be ergodic when it reaches an absorbing state consisting of a single strategy.

Figure 1 illustrates our model by showing how a population of the above-mentioned two strategies, i.e., 27 and 255, evolves on a one-dimensional ring with \( L = 8 \) cells. Both the environmental parameter \( \alpha \) and the mutation probability \( \mu \) are set to be zero to help follow the rules in a fully deterministic way. Note that all the cells are updated in parallel as \( t \) increases by one in this example, and this will also be the case of our Monte Carlo calculation in the next section (see Ref. 24 for possible effects of update rules on time evolution). However, the long-time behavior presented below shows no significant difference even when we use a random asynchronous update rule.

III. RESULT

Let us now include all the 256 strategies of \( S \) and simulate the model on a larger ring structure with \( L = 64 \) cells (Fig. 2). Initially at \( t = 0 \), every cell is occupied by an organism with a randomly drawn strategy from \( S \). The colors represent strategy indices from 0 to 255 (see the color box on the right), and the white cells are empty.
FIG. 3: Frequencies of strategies under $\alpha = 0.1$. The system size is $L = 1024$, the mutation rate is $\mu = 10^{-3}$, and all the results are averaged over $M = 10^3$ independent realizations. The colors represent strategy indices from 0 to 255 as in Fig. 2. Initially at $t = 0$, every strategy starts with an equal frequency, but selection favors strategy 27 and its variants.

Reddish strategies take over when $\alpha = 0.6$ or 0.8, but their cluster is porous, and the porosity will gradually vanish as $\alpha \to 1$.

To quantify the behavior, we calculate the ensemble-averaged frequency of strategy $k$ as follows:

$$f_k(t) = \frac{1}{M} \sum_{m=1}^{M} \frac{N_k^{(m)}(t)}{L},$$

where $M$ is the number of independent Monte Carlo realizations and $N_k^{(m)}(t)$ is the number of organisms playing strategy $k$ at time $t$ in the $m$th realization (Fig. 3). To obtain its value in a steady state, we remove transient behavior for a certain initial period $T$ and then take an average over $P$ generations:

$$\phi_k = \frac{1}{P} \sum_{t=T+1}^{T+P} f_k(t).$$

The total density of population,

$$\rho = \sum_{k \in S} \phi_k,$$

is a measure of collective interests for this group of organisms.

Let us check how these observables behave as $\alpha$ varies. Figure 4 shows that an unfavorable environment with small $\alpha$ tends to favor bluish nonaggressive strategies such as 27, and they are replaced by more and more aggressive ones as $\alpha$ increases, which is entirely consistent with Fig. 2. Note that strategies 31 and 59 exhibit identical behavior because they are related by left-right symmetry, and the same statement holds between 127 and 191. An interesting point is that the total density of the population $\rho$ decreases as the environment becomes more and more favorable between 0.2 and 0.5.

IV. DISCUSSION

To illustrate the basic picture, it is instructive to work with a reduced set of strategies. We choose 27 and 255, the most favored ones for small and large $\alpha$’s, respectively (Fig. 4). The former strategy is able to increase the population density $\rho$ up to 100% by replicating itself in a nonaggressive way. Thus, it may be called a
“cooperating” strategy. The latter strategy is the most aggressive one, and we may call it a “defecting” strategy.

Let us consider three consecutive cells, each of which is either C (cooperating), D (defecting), or E (empty). From the configuration of these three cells, we can discuss the replication dynamics in the middle cell. Let \( \eta^+_C \) be the rate for an empty cell to be occupied by a cooperator. By enumerating all the possible cases, we see that

\[
\eta^+_C = (\phi_{ECE} + \phi_{CCE}) + (\phi_{CDE} + \phi_{DEC}) \frac{\alpha}{2} + \phi_{CC} \alpha, \tag{6}
\]

where \( \phi_{XYZ} \) is the frequency for the three consecutive cells to have states X, Y, and Z, respectively. On the other hand, a C cell becomes either a D cell or an E cell with a rate of

\[
\eta^-_C = (\phi_{ECD} + \phi_{CCE} + \phi_{DCE} + \phi_{DCC}) \left(1 - \frac{\alpha}{2}\right) + \phi_{DDC} \left(1 - \frac{\alpha^2}{3}\right). \tag{7}
\]

Similarly, we can write the rates \( \eta^+_D \) for creation or annihilation of D cells. However, to know \( \phi_{XYZ} \), the statistics of five consecutive cells is required, and this hierarchy generally goes on \textit{ad infinitum}. As an approximation, we will factorize \( \phi_{XYZ} \) into \( \phi_X \phi_Y \phi_Z \), where \( \phi_X \) denotes the frequency of X cells [see Eq. (4)], and this mean-field approximation is valid in the absence of spatial correlations. Figure 2(f) suggests that D cells form a homogeneous mixture with E cells (without strong spatial correlations) for large \( \alpha \) in the steady states. We can estimate the frequency of D cells using a mean-field approximation in such \((D + E)\) clusters. In these clusters, \( \eta^+_D \) is given by

\[
\eta^+_D = (\phi_{EDD} + \phi_{DDE} + \phi_{DED}) + \phi_{DED} \alpha, \tag{8}
\]

\[
\eta^-_D = (\phi_{EED} + \phi_{EDE}) \left(1 - \alpha\right) + \phi_{DDE} \left(1 - \alpha^2\right). \tag{9}
\]

Equating \( \eta^+_D \) and \( \eta^-_D \), the equilibrium frequency of D cells in a \((D + E)\) cluster is obtained as

\[
\phi_D = \frac{6 - 3\alpha - \sqrt{12 - 12\alpha + \alpha^2}}{2(3 - 3\alpha + \alpha^2)} = 1 + 2\epsilon + 7\epsilon^2 + 35\epsilon^3 + \ldots, \tag{10}
\]

where we used \( \phi_E = 1 - \phi_D \) and set \( \epsilon \equiv \alpha - 1 < 0 \) (Fig. 5). Equation (10) agrees well with our numerical results for large \( \alpha \), implying that the whole system can be described as a single \((D + E)\) cluster. On the other hand, the system is mostly filled with C cells for small \( \alpha \). This suggests existence of a transition between C and \((D + E)\) phases at a certain threshold \( \alpha = \alpha^* \).

To estimate \( \alpha^* \), let us assume that a \((D + E)\) cluster has an interface with a C cluster. The most probable situation for growth of the C cluster is found when the two nearest cells to the interface on the \((D + E)\) side are empty. The simplest estimate for this probability would be \( \phi_{EEZ}^2 = (1 - \phi_D)^2 \), under the assumption that the bulk behavior inside a \((D + E)\) cluster is mostly valid even in the vicinity of the interface. The second contribution is given by another configuration in which D and C compete for an empty cell in the middle, and this contributes \( \phi_E(1 - \phi_D)(\alpha/2) \) because C wins with probability \( \alpha/2 \). On the other hand, the \((D + E)\) cluster can proceed by one cell with probability \( \phi_D \times (\alpha/2) \) because the front must be filled with \( D \) and the invasion succeeds with probability \( \alpha/2 \). If we compare these two events, then the latter becomes more probable for large \( \alpha \), and the threshold value is estimated by equating them, i.e.,

\[
(1 - \phi_D)^2 + \phi_D(1 - \phi_D)(\alpha/2) = \phi_D \times (\alpha/2), \tag{12}
\]

which, together with Eq. (10), results in \( \alpha^* \approx 0.64 \) (Fig. 5). We note that this may well be an overestimate because the actual frequency of \( D \) is likely to be higher than predicted by Eq. (10) near the interface, where the competition between C and D would be less intense than between two D’s.

The above mean-field calculation can be modified by using the pair approximation [27], according to which three-point and four-point correlation functions are approximated as

\[
\phi_{XYZ} \approx \frac{\phi_{XY} \phi_{YZ}}{\phi_Y}, \tag{13}
\]

and

\[
\phi_{XYZW} \approx \frac{\phi_{XY} \phi_{YZ} \phi_{ZW}}{\phi_Y \phi_Z}, \tag{14}
\]

respectively (see Refs. [28,29] for further modification beyond the pair approximation). If we deal with a \((D + E)\) cluster, then we need five correlation functions, i.e., \( \phi_D, \phi_E, \phi_{DD}, \phi_{DE} = \phi_{ED}, \phi_{EE} \), but only two of them are independent because \( \phi_D = 1 - \phi_E = \phi_{DD} + \phi_{DE} = 1 - (\phi_E + \phi_{ED}) \). If we find \( \phi_D \) and \( \phi_{DD} \), for example, then the other three are determined by these relations. Regarding \( \phi_{DE} \), the rates of creating and annihilating \( DD \) cells are given as

\[
\eta^+_D = (\phi_{EDE} + \phi_{DEE} + \phi_{EDE}) + \alpha(\phi_{EDE} + \phi_{DEE} + \phi_{DEE} + \phi_{DE}) + \alpha^2(\phi_{EDE} + \phi_{DEE}), \tag{15}
\]

and

\[
\eta^-_D = (1 - \alpha^4) \phi_{DDE} + (1 - \alpha^4) \phi_{DDD} + (1 - \alpha^2) \phi_{DDE}, \tag{16}
\]

respectively. By solving \( \eta^+_D = \eta^-_E \) and \( \eta^+_E = \eta^-_D \) with the pair approximation [Eqs. (13), (15), and (16)], we obtain \( \phi_D \) as a function of \( \alpha \) which is shown as the dash-dotted curve in Fig. 5. Although its explicit expression is not illuminating, a few points are worth mentioning: First, the pair-approximated version of \( \phi_D \) has the same Taylor series to the order of \( \epsilon^3 \) as given by the mean-field
calculation [see Eq. (11)]. Second, if we also write $\phi_{DD}$ as a function of $\alpha$, then we obtain the connected correlation function $\tilde{\phi}_{DD} \equiv \phi_{DD} - \phi_{D} = -4\delta^{3} + \ldots$, which is indeed small and thus consistent with the mean-field-like ideas behind our approximate calculation. Third, the system has four solution branches, and the physical solution, having both $\phi_{D}$ and $\phi_{DD}$ inside the unit interval, changes its branch at $\alpha \approx 0.60$, which might indicate an improved estimate of $\alpha^{*}$.

To understand the nonequilibrium phase transition between $C$ and $D$ more precisely $[32, 33]$, we conduct Monte Carlo simulation and observe the following quantity: Let $P_{C}(t)$ be the probability to have at least one $C$ cell at time $t$ when the simulation started at $t = 0$ with a single $C$ cell in a system filled with $D$. The result in Fig. 5(a) shows that it decays as $P_{C}(t) \sim t^{-\delta}$ at $\alpha^{*} = 0.4991 \pm 0.0001$. From $4 \times 10^{5}$ samples for each $\alpha$, we estimate the decay exponent as $\delta = 0.50 \pm 0.02$, where the error mainly originates from the uncertainty in $\alpha^{*}$. The number of $C$ cells is another quantity expected to show power-law behavior $N_{C}(t) \sim t^{\theta}$ [Fig. 5(b)], and we estimate the exponent as $\theta = -0.01 \pm 0.03$. We have also obtained consistent results by exchanging $C$ and $D$ in the initial configuration (not shown).

To conclude, our approximate calculation predicts that $C$ will densely occupy the whole system if $\alpha < \alpha^{*}$. Otherwise, the system will be occupied by a mixture of $D$ and $E$, among which the fraction of $D$ is described by Eq. (10). The total density of population should decrease as $\alpha$ exceeds $\alpha^{*}$ because $\phi_{D}(\alpha = \alpha^{*})$ is far smaller than 100%. Our numerical results suggest that the behavior at $\alpha = \alpha^{*}$ can be described by random walks of domain walls because the survival probability behaves as $P_{C}(t) \sim t^{-\delta}$ with $\delta \approx 1/2$ and the average number of $C$ cells is approximately constant.

V. SUMMARY

To summarize, we have studied an evolutionary game in which replication strategies are inherited by the next generation and the survival probability in competition depends on neighbors’ strategies as well as one’s own. We have examined evolution of the population with varying the environmental favorability that determines the chance of surviving competition. Our finding is that the population sometimes flourishes better when the survival probability is smaller because it eventually evolves to a more cooperative strategy. Although we have focused on a one-dimensional system to see the effects of spatiality most clearly, it is entirely plausible that the effects will diminish in higher dimensions and disappear in a well-mixed population. Exact identification of the critical dimension is left as a future work.

A common assumption in microeconomics is that production functions monotonically increase in all inputs so that output quantities do not decrease when any input quantity is increased. Our result suggests that the monotonicity assumption may not always hold when an evolutionary process is involved, if we regard $\alpha$ as a measure of input resources and the population density $\rho$ as the output. If the organisms under consideration are coupled with the input resources through a predator-prey interaction, then it implies that the coupling will be described as nonlinear, as opposed to the linear coupling in the Lotka-Volterra type, due to the intraspecific interaction among different behavioral strategies. More specifically, the mean-field analysis discussed above shows that assortative clustering can result in nonmonotonic behavior through interfacial dynamics between two competing clusters. It demonstrates the role of assortative clustering in evolution of cooperation under the condition of resource scarcity.

Acknowledgments

S.K.B. was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant No. NRF-2020R1I1A2071670). H.C.J. was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant No. NRF-2018R1D1A1A02086101).
