Spontaneous Group Formation in the Seceder Model

Peter Dittrich[1], Fredrik Liljeros[2], Arne Soulier[3], and Wolfgang Banzhaf[1]

[1] *University of Dortmund, Dept. of Computer Science XI, D-44221 Dortmund*

[2] *Stockholm University, Dept. of Sociology, S-10691 Stockholm*

[3] *Columbia University, Dept. of Physics, New York, NY 10027*

(Submitted: November 10, 1999; Compiled: December 7, 1999)

Abstract

The seceder model shows how the local tendency to be different gives rise to the formation of groups. The model consists of a population of simple entities which reproduce and die. In a single reproduction event three individuals are chosen randomly and the individual which possesses the largest distance to their center is reproduced by creating a mutated offspring. The offspring replaces a randomly chosen individual of the population. The paper demonstrates the complex group formation behavior and its dependency on the population size.

PACS numbers: 05.65.+b, 87.23.Kg, 87.23.Cc, 87.23.Ge,

Typeset using REVTEX

The question of how groups emerge spontaneously from local interactions of individuals is investigated in many different disciplines like biology, sociology or computer science. There has also been an increasing interest from statistical physics to deal with simple evolutionary models [1–3]. In evolutionary biology the question how evolutionary branching and speciation take place is for example approached by developing formal models which demonstrate the formation of groups [2–6]. These models are individual-based in contrast to macro-evolution models which assume a species or group as a given elementary unit [1,7,8]. The diffusion and separation of individuals in genotype or trait space is either achieved by drift in a neutral fitness landscape [4] or by introducing an explicit fitness function [2,3,9] which causes disruptive selection. Sometimes additional explicit functions are introduced to model strength of competition between individuals and ecological interactions [2]. Such functions are also needed to model the benefit of communication among groups on several levels [3].

The seceder model developed in this letter is a microscopic model of an evolving population where the fitness landscape depends on the current population structure, like in [2,4,10]. The proposed mechanism is simple compared to other individual-based models [3,11,12] for the formation of species or hierarchical organizations. But despite of its simplicity it shows comparably complex behavior. The seceder model does not require global energy functions [2,3], spatially separated populations [5,13] or sexual recombination [2,4,9].

The question of how microlevel actions explain macrolevel regularities is also a central question in sociology [14–16], Here, the seceder model may be a contribution as a social mechanism [15] for explaining how individual imitative behavior for the purpose of being different counter-intuitively can lead to the formation of groups on the macrolevel. In this context, the seceder model could be seen as reflecting a choice situation where individuals collect a bounded amount of information by observing other individual's behavior and, then choose the most unique alternative. In this way, the model could be used as a description for how groups with different dress codes emerge in populations of individuals, as a consequence of the individuals' need to express their individuality [17].

It should also be noted that the mechanism of the seceder model can be used to build practical applications in computer science. For example, it can be used as a diversity maintenance method for evolutionary optimization algorithms [18] where the reduction of diversity often causes a premature convergence and thus a bad performance of the optimization algorithm [19].

In the seceder model the formation of groups is an emergent process based on the following simple local third-order collision rule: Three individuals are chosen randomly from the population and the individual which has the largest distance in a predefined metric to the other two is reproduced by creating a mutated copy (offspring). More formally, the seceder model is defined as follows: Each *individual* is represented by a real number. The *population* of size M is represented by an array (or multi set) $P = \{p_1, \dots, p_M\}$ of individuals $p_i \in \mathcal{R}$. The population evolves over time according to the following algorithm: (1) Three individuals s_1, s_2, s_3 are chosen randomly from the population without removing them. (2) One individual $\mu = f_{sel}(s_1, s_2, s_3)$ is selected as a parent according to the selection function

$$f_{sel}(s_1, s_2, s_3) = \begin{cases} s_1 & : \quad F_1 \geq F_2 \wedge F_1 \geq F_3 \\ s_2 & : \quad F_2 \geq F_1 \wedge F_2 \geq F_3 \\ s_3 & : \quad \text{otherwise} \end{cases} \tag{1}$$

with $F_i = |s_i - \frac{1}{3}(s_1 + s_2 + s_3)|$. (3) The offspring $\lambda = \mu + N(0,1)$ is created by adding a normally distributed random number with mean 0 and variance $\sigma = 1$ to the parent μ. (4) The offspring λ replaces a randomly chosen individual of the population.

One iteration of the above algorithm is called a *step*. M iterations are called a *generation* which is used to measure time [12]. *Distance* between two individuals is measured by the Euclidean distance (see Def. of F_i). *Mutation* is performed by adding a normally distributed random number as in evolution strategies [18]. For our experiments the population is initialized with copies of one genotype, $p_i = 0$ at $t = 0$. The algorithm implies that the population size is constant and that an individual may have an arbitrary number of offsprings including no offspring.

To visualize how the *population structure* evolves over time we plot individuals of an evolving population as dots in a 2-D plane (Fig. 1 and 4). An individual's value (genotype) specifies the ordinate of the dot and the time the abscissa. The *diameter* of a population P is the largest distance between individuals that can be found in P.

We also measure the number of groups. Here, a *group* is defined as a subset of the population which is separated from the rest of the population by at least a distance equal to d, called *gap size*. The *number of groups* is then computed by partitioning the population P so that (1) each element in P is assigned to exactly one partition, (2) the distance between two object belonging to different partitions is at least d, and (3) two elements in a partition are connected by a sequence of elements belonging to the same partition with the distance between two successive elements in that sequence being smaller than d. In other words, a partition or group can not be subdivided into two subgroups which are separated by at least a gap size d.

We begin our discussion with the time evolution of the population structure in the transient phase when the population is initialized at $t = 0$ with copies of the same individual $p_i = 0$. Figure 1 shows that at the outset the population spreads out quickly and increases its diameter. For population sizes $M > 100$ a complex group formation pattern appears. New groups are formed by diversification (splitting up) of existing groups. Existing groups may also vanish (go extinct). For very small population sizes ($M < 10$) group formation cannot be observed. For population sizes around 20 two groups are formed only temporarily. The qualitative behavior of the time evolution of the population diameter also depends on the population size (Fig. 1 and 2). For small population sizes the diameter increases first and then saturates. It fluctuates heavily because groups at the population rim vanish. For large population sizes the diameter seems to increase for ever with linear speed. Because of the existence of random death the upper or the lower arm may die out with a very low probability. In such a case the diameter shrinks suddenly again. In a large population the probability that one main arm dies out is small enough to be neglected.

Figure 3 shows how the number of groups depends on the population size. Average values of many runs are shown for each population size and for three different gap sizes

$d = 10\sigma, 20\sigma$, and 30σ (here, mutation strength $\sigma = 1$). The figure supports the impression gained from single runs (like shown in Fig. 1) that there is a non-trivial relation ship between the population size and the qualitative behavior of group formation. Two "steps" can be clearly observed for all three gap sizes. The question whether for $M > 400$ another step appears can not be answered yet. For this purpose a new, computationally more efficient method for measuring the number of groups has to be developed.

In order to illustrate the long-term behavior of the seceder model Fig. 4 shows instances of four runs with population size $M = 20, 200, 2000$, and 20000. The simulation time is 50 times longer than in Fig. 1. On this scale groups cannot be distinguished when the population is small (here, $M = 20$) because the groups are too close and too short-lived. The time evolution appears to be a random walk through genotype space. For large population sizes (here, $M = 2000$, and 20000) stable main arms form. The typical pattern is three main arms where the upper and lower arm disperse and the middle arm stays close to the center of the population. The run with $M = 200$ illustrates that a main arm may die out by chance and that new arms can appear. Looking at Fig. 1 and 4 it seems that the evolution pattern of the population structure is fractal. That can for example be seen the cusps like structures formed in Fig. 1 in the $M = 20000$ population.

In next paragraph an ordinary differential equation (ODE) model is introduced for the special case of a population with only a few very different genotypes. The ODE model assumes that the offspring is equal to one of the three individuals chosen in one step. The resulting ODE model is similar to the hypercycle model [10], the catalytic network equation [20], and the replicator equation [21].

Assume that the population represented by the multi set $P = \{p_1, p_2, \ldots, p_M\}$ consists of n different genotypes given by the set $G = \{g_1, g_2, \ldots, g_n\}, n \ll M$ such that $\forall p \in P : p \in G$. We further assume that the distance between two genotypes $g_i, g_j \in G, i \neq j$ is large, that the mutation rate σ is small (e.g. $\sigma \to 0$), and that the population size M is large (e.g. $M \to \infty$). Then the state of the population can be represented by a point $x = (x_1, x_2, \ldots, x_n)$ in the concentration simplex $S_n = \{(x_1, x_2, \ldots, x_n) | \sum_{i=1}^{n} x_i = 1, x_i \geq 0\}$ where x_i is the concentration of genotype g_i in P. The dynamics of the model is given by the following ODE:

$$\dot{x}_l = \sum_{i,j,k=1}^{n} \alpha^l_{i,j,k} x_i x_j x_k - x_l \text{ for } l \in \{1, 2, \ldots, n-1\}, \tag{2}$$

$$x_n = 1 - \sum_{i=1}^{n-1} x_i. \tag{3}$$

The rate constants $\alpha^l_{i,j,k}$ are defined by:

$$\alpha^l_{i,j,k} = \begin{cases} 1 & : \quad g_l = f_{sel}(g_i, g_j, g_k), \\ 0 & : \quad \text{otherwise.} \end{cases} \tag{4}$$

Note that for the seceder model the dilution flux $\Phi(x) = \sum_{i,j,k,l=1}^{n} \alpha^l_{i,j,k} x_i x_j x_k = 1$ is equal to one because at each step exactly one individual is removed from the population. The ODE system has been derived by interpreting the seceder model as a chemical reaction system where three molecules g_i, g_j, g_k collide in order to catalyze the formation of a fourth molecule g_l. The reaction is of third order and can be written as[1]: $g_i + g_j + g_k \xrightarrow{\alpha^l_{ijk}} g_i + g_j + g_k + g_l$. Usual mass action kinetics and consideration of the dilution flux leads to Eq. 2. Because g_l is equal to g_i, g_j, or g_k the reaction is a replication [10,21].

To demonstrate the ODE model we consider the following example where the population consists of only $n = 3$ "symmetric" genotypes: $G = \{g_1, g_2, g_3\} = \{g_1, 0, -g_1\} = \{10^5, 0, -10^5\}$. A similar situation is shown in Fig. 4 for $M = 2000$ (bottom, left) at $t = 6000$. The ODE system $\dot{x} = (\dot{x}_1, \dot{x}_2)^T = F(x_1, x_2, x_3)$ is defined by:

$$F(x_1, x_2, x_3) = \begin{pmatrix} x_1^3 + 3x_1x_2^2 + 3x_1x_3^2 + 3x_1x_2x_3 & -x_1 \\ x_2^3 + 3x_2x_1^2 + 3x_2x_3^2 & -x_2 \end{pmatrix} \tag{5}$$

and $x_3 = 1 - x_1 - x_2$. For a fixed point analysis we set $F(x) \overset{!}{=} 0$ and obtain the following fixed points $x \in \{(1,0), (0,1), (0,0), (0.5, 0.5), (0.5, 0), (0, 0.5), (0.4, 0.2)\}$. For a stability analysis (which assumes a perturbation with radius ϵ around a fixed point) we derive

---

[1] To be more precise the reaction rule should be balanced like $g_i + g_j + g_k + X \longrightarrow g_i + g_j + g_k + g_l$, where X is an implicit substrate which concentration is constant.

the Jacobian of F and calculate for every fixed point x the eigen values of the Jacobian: $EV(J(F)(x)) = \{(2,2),(2,2),(2,2),(1.25,-1),(-1,0.5),(1.25,-1),(-0.88,-0.4)\}$ So, only the ratio $(0.4, 0.2, 0.4)$ is an asymptotic stable fixed point. Figure 5 shows that this fixed point also describes the concentration relation in the seceder model adequately. For the other fixed points the ODE model predicts correctly the concentration relation of the genotypes, but does not correctly predict the stability. In the ODE model these fixed points are unstable. A small perturbation would cause the system (ODE model) to run into the stable fixed point attractor $(0.4, 0.2, 0.4)$. In the seceder model this is impossible because a new genotype (e.g. g_1) cannot suddenly appear. In conclusion, the simple ODE model introduced here could explain the qualitative dynamics for the conditions given above for the interior of the concentration simplex S_n. But it cannot explain the qualitative dynamics on and close to the border of the concentration simplex S_n where discrete population size become visible effects.

In summary the seceder model shows how a simple local mechanism – the advantage of being different – gives rise to complex group formation phenomena. Can the seceder model be used to explain (at least partly) the formation and evolution of groups in a biological and social context ? We think that it is a promising candidate (partly) for explaining group or sympatric species formation [22] in an evolutionary context for the following reason: 1) The selection function f_{sel} represents the advantage which stems from occupying a lowly populated niche. 2) The macroscopic pattern of the evolving population structure is similar to patterns derived from the fossil record. 3) The model can be easily extended to include: bounded genotype space, environmental stress [7], sexual recombination, external fitness pressure, a spatial world, higher-dimensional genotypes, etc. 4) Compared to macroevolutionary models which assume a species to be a given entity [1,7], the seceder model can explain group formation and evolution based on local interactions of individuals.

We wish to thank the SFI complex system summer school for providing the stimulating interdisciplinary atmosphere where the idea for this work has been originated. And we wish to thank P. Hedström and T. Halpin-Healy for helpful comments. W.B. and P.D.

acknowledge support from the DFG (Ba 1042/2-2). F.L. research has been partially financed with grants from STINT (97/1837) and HSFR (F0688/97). C++ source code for the seceder model is available from **http://ls11-www.cs.uni-dortmund.de/alife/seceder**.

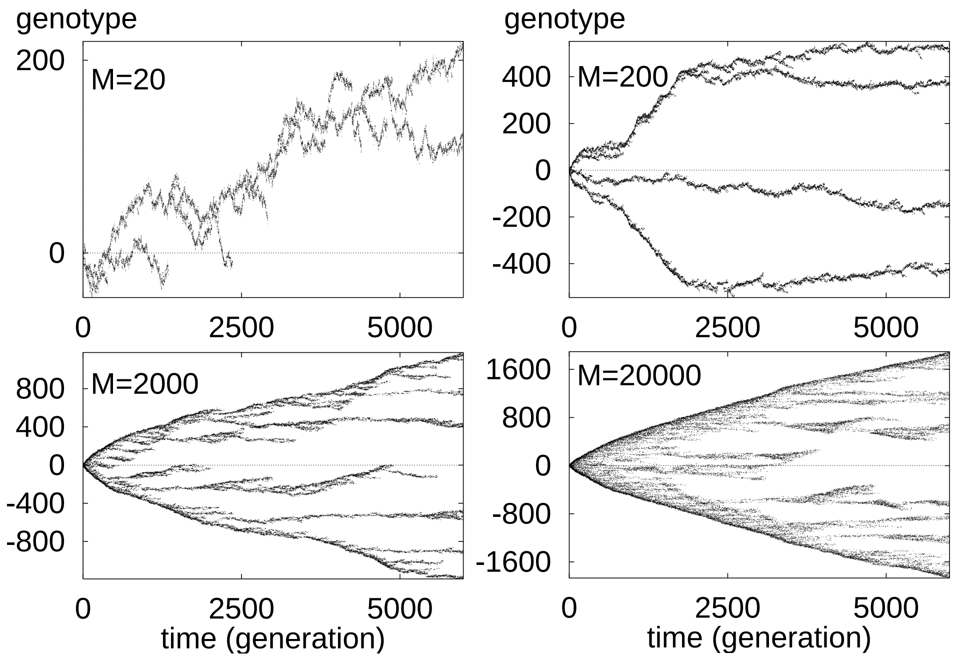

FIG. 1. Typical evolution of the population structure for population size $M = 20, 200, 2000, 20000$ of the seceder model. The population is initialized at $t = 0$ with $p_i = 0$. At each time (generation) every individual of the population is plotted as a small dot with its genotype value specifying its ordinate value (vertical position).

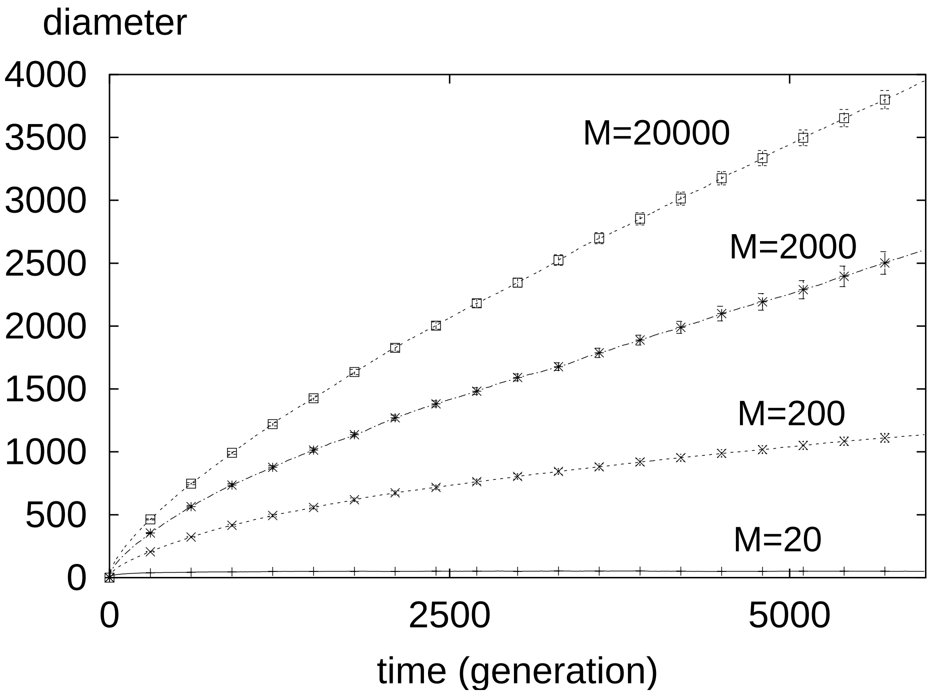

FIG. 2. Change in the average diameter over time for population size $M = 20, 200, 2000$, and 20000. The average is taken over 1000,200,20, and 10 runs for each population size, respectively.

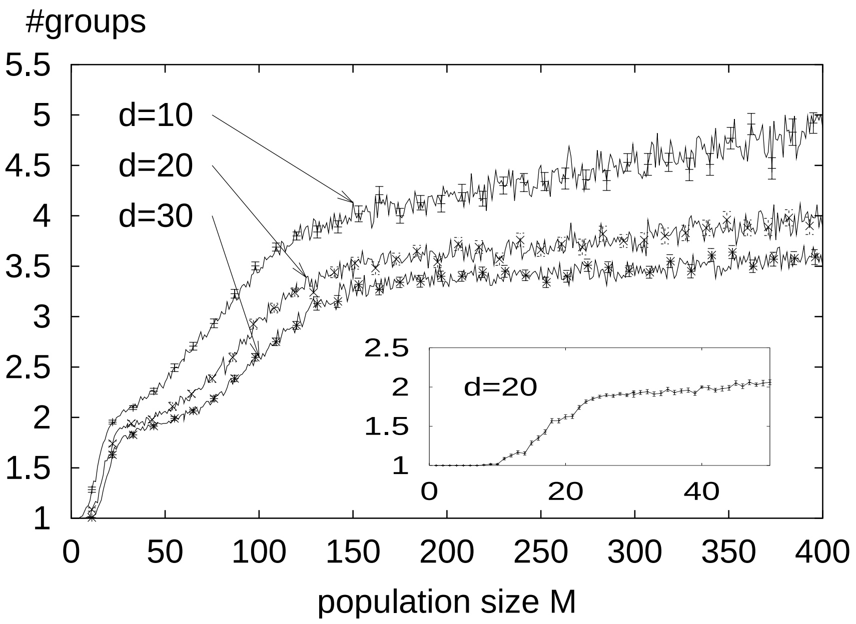

FIG. 3. Average number of groups measured at $t = 2000$ for different population sizes and three different gap sizes $d = 10\sigma, 20\sigma$, and 30σ. Average taken over 300 runs (for $M < 120$) and 100 runs (for $M > 120$) for each population size M. Population initialized with $p_i = 0$ at $t = 0$. Error bars indicating the standard error are plotted for a quarter of measured points.

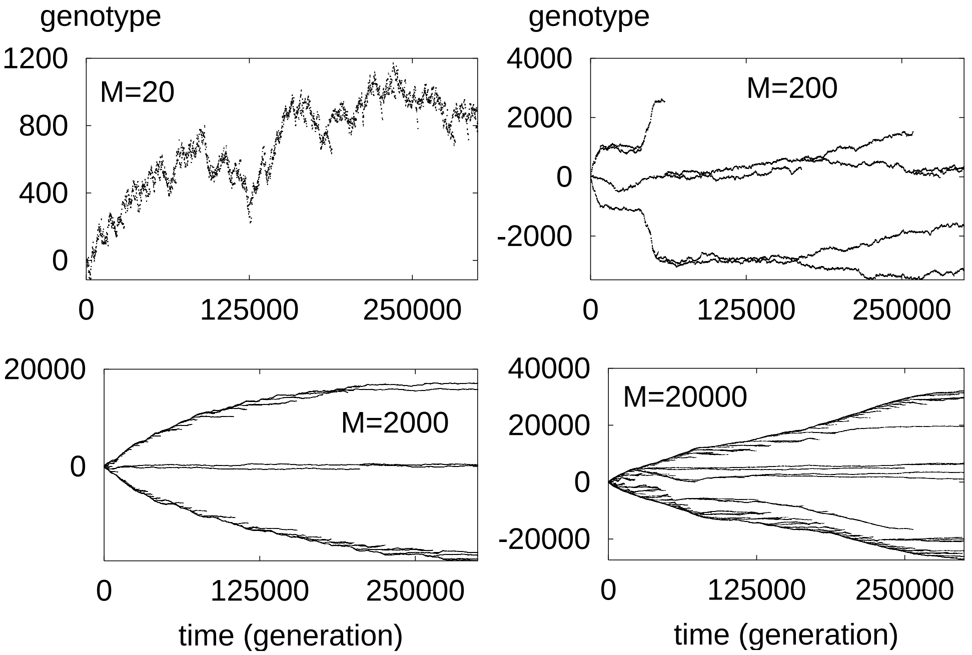

FIG. 4. Typical long-term evolution of the population structure for population size $M = 20, 300, 2000, 20000$. For smaller populations (e.g. M=200) which are beyond the stage where they can form stable groups sometimes do not show the full almost symmetric boundaries as the larger populations (M=2000, M=20000) do. They are still in a stage of reaching this final structure and therefore can have "odd" behavior, e.g. a number of representations not grow in population diameter linearly but in stages, or sidearms can die out and therefore the population diameter shrinks drastically.

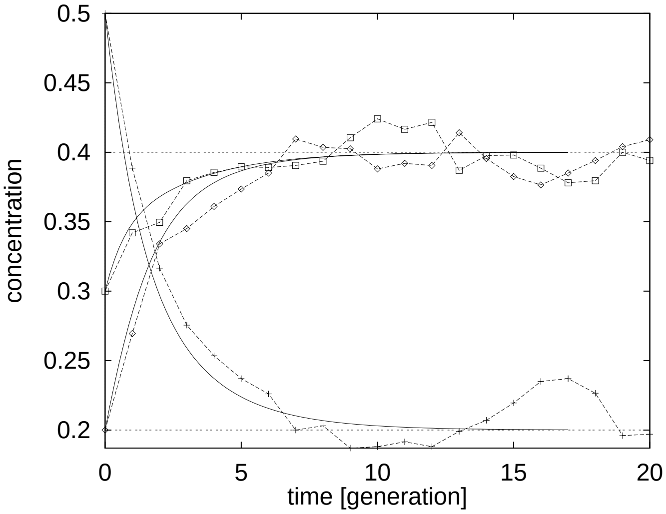

FIG. 5. Comparison of the seceder model (dotted lines) with its ODE model (solid lines) for population size $M = 2000$. In the seceder model the population is initialized with 600, 1000, and 400 copies of the genotypes $g_1 = 10^5, g_2 = 0$, and $g_3 = -10^5$, respectively. Thus the initial condition for the ODE model is set to $x_1(t = 0) = 0.3, x_2(t = 0) = 0.5$, and $x_3(t = 0) = 0.2$. The figure shows that the seceder model and the ODE converge to the ratio of the asymptotic stable fixed point at at $(0.2, 0.4, 0.2)$.

REFERENCES

[1] P. Bak and K. Sneppen, Phys. Rev. Lett. **71**, 4083 (1993).

[2] U. Dieckmann and M. Doebeli, Nature **400**, 354 (1999).

[3] B. Drossel, Phys. Rev. Lett. **82**, 5144 (1999).

[4] P. Higgs and B. Derrida, J. Phys. A **24**, 985 (1991).

[5] F. Manzo and L. Peliti, J. Phys. A **27**, 7079 (1994).

[6] G. F. Turner and M. T. Burrows, Proc. R. Soc. Lond. Ser. B **260**, 287 (1995).

[7] M. E. J. Newman, Proc. R. Soc. London Ser. B **263**, 1605 (1996).

[8] C. Wilke and T. Martinetz, Phys. Rev. E **59**, R2512 (1999).

[9] A. S. Kondrashov and F. A. Kondrashov, Nature **400**, 351 (1999).

[10] M. Eigen and P. Schuster, Naturwissenschaften **64**, 541 (1977).

[11] F. Bagnoli and M. Bezzi, Int. J. Mod. Phys. C **9**, 555 (1998).

[12] P. Dittrich and W. Banzhaf, Artificial Life **4**, 203 (1998).

[13] K. Johst, M. Doebeli, and R. Brandl, Proc. R. Soc. Lond. Ser. B **266**, 1147 (1999).

[14] T. T. Schelling, *Micro Motives and Macrobehavior* (Norton, New York, 1978).

[15] *Social Mechanisms : An Analytical Approach to Social Theory*, edited by P. Hedström and R. Swedberg (Cambridge Univ. Pr., Cambridge, MA, 1998).

[16] N. Saam, Sociol. Methol. **29**, 43 (1999).

[17] P. Bourdieu, *Distinction: A Social Critique for the Judgement of Taste* (Routledge and Kegan Paul, London, 1984).

[18] *Handbook of Evolutionary Computation*, edited by T. Bäck, D. B. Fogel, and Z. Michalewicz (IOP Publishing and Oxford University Press, Bristol, UK, 1997).

[19] N. Shamir, D. Saad, and E. Marom, Complex Systems **7**, 327 (1993).

[20] P. F. Stadler, W. Fontana, and J. H. Miller, Physica D **63**, 378 (1993).

[21] J. Hofbauer and K. Sigmund, *Dynamical Systems and the Theory of Evolution* (University Press, Cambridge, UK, 1988).

[22] J. Maynard Smith, Am. Nat. **100**, 637 (1966).