A LETTERS JOURNAL EXPLORING THE FRONTIERS OF PHYSICS

OFFPRINT

Solvable model for template coexistence in protocells

J. F. FONTANARI and M. SERVA

EPL, **101** (2013) 38006

epl

A Letters Journal Exploring the Frontiers of Physics

AN INVITATION TO SUBMIT YOUR WORK

www.epljournal.org

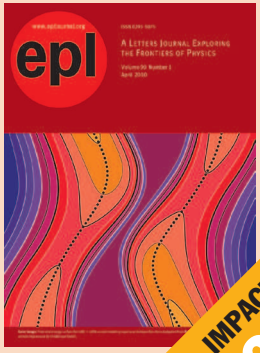
The Editorial Board invites you to submit your letters to EPL

EPL is a leading international journal publishing original, high-quality Letters in all areas of physics, ranging from condensed matter topics and interdisciplinary research to astrophysics, geophysics, plasma and fusion sciences, including those with application potential.

The high profile of the journal combined with the excellent scientific quality of the articles continue to ensure EPL is an essential resource for its worldwide audience. EPL offers authors global visibility and a great opportunity to share their work with others across the whole of the physics community.

Run by active scientists, for scientists

EPL is reviewed by scientists for scientists, to serve and support the international scientific community. The Editorial Board is a team of active research scientists with an expert understanding of the needs of both authors and researchers.



IMPACT FACTOR 2.753*

*As ranked by ISI 2010

IMPACT FACTOR
2.753*
* As listed in the ISI® 2010 Science Citation Index Journal Citation Reports

OVER
500 000
full text downloads in 2010

30 DAYS
average receipt to online publication in 2010

16 961
citations in 2010
37% increase from 2007

"We've had a very positive experience with EPL, and not only on this occasion. The fact that one can identify an appropriate editor, and the editor is an active scientist in the field, makes a huge difference."

Dr. Ivar Martinv
Los Alamos National Laboratory, USA

Six good reasons to publish with EPL

We want to work with you to help gain recognition for your high-quality work through worldwide visibility and high citations.

1. **Quality** – The 40+ Co-Editors, who are experts in their fields, oversee the entire peer-review process, from selection of the referees to making all final acceptance decisions
2. **Impact Factor** – The 2010 Impact Factor is 2.753; your work will be in the right place to be cited by your peers
3. **Speed of processing** – We aim to provide you with a quick and efficient service; the median time from acceptance to online publication is 30 days
4. **High visibility** – All articles are free to read for 30 days from online publication date
5. **International reach** – Over 2,000 institutions have access to EPL, enabling your work to be read by your peers in 100 countries
6. **Open Access** – Articles are offered open access for a one-off author payment

Details on preparing, submitting and tracking the progress of your manuscript from submission to acceptance are available on the EPL submission website **www.epletters.net**.

If you would like further information about our author service or EPL in general, please visit **www.epljournal.org** or e-mail us at **info@epljournal.org**.

EPL is published in partnership with:

European Physical Society | Società Italiana di Fisica | EDP Sciences | IOP Publishing

A Letters Journal Exploring the Frontiers of Physics

February 2013

EPL, **101** (2013) 38006
doi: 10.1209/0295-5075/101/38006

www.epljournal.org

Solvable model for template coexistence in protocells

J. F. Fontanari[1] and M. Serva[2(a)]

[1] *Instituto de Física de São Carlos, Universidade de São Paulo - Caixa Postal 369, 13560-970 São Carlos, SP, Brazil*
[2] *Departamento de Biofísica e Farmacologia, Universidade Federal do Rio Grande do Norte 59072-970 Natal, RN, Brazil*

received 11 October 2012; accepted in final form 24 January 2013
published online 14 February 2013

PACS 87.10.Ed – Ordinary differential equations (ODE), partial differential equations (PDE), integrodifferential models
PACS 89.75.Fb – Structures and organization in complex systems
PACS 05.70.Jk – Critical point phenomena

Abstract – Compartmentalization of self-replicating molecules (templates) in protocells is a necessary step towards the evolution of modern cells. However, coexistence between distinct template types inside a protocell can be achieved only if there is a selective pressure favoring protocells with a mixed template composition. Here we study analytically a group selection model for the coexistence between two template types using the diffusion approximation of population genetics. The model combines competition at the template and protocell levels as well as genetic drift inside protocells. At the steady state, we find a continuous phase transition separating the coexistence and segregation regimes, with the order parameter vanishing linearly with the distance to the critical point. In addition, we derive explicit analytical expressions for the critical steady-state probability density of protocell compositions.

Introduction. – Explaining the coexistence among selfish individuals is an alluring issue that pervades all disciplines concerned with the emergence and stability of complex structures evolving under the rules of natural selection. A series of classical quandaries can be traced back to this issue such as the plankton paradox in ecology [1], the tragedy of commons in sociology [2] and the information crisis of prebiotic evolution [3], just to name a few. The many ingenious solutions proposed to tackle those dilemmas, *e.g.*, coexistence in flows with chaotic mixing [4], kin selection [5], reciprocal altruism [6], cyclic cooperative interactions [7] and group selection [8], have become major research topics by themselves.

Here we revisit the problem of coexistence of selfish individuals using the group selection framework in the context of prebiotic evolution [9–12]. We recall that, in this context, coexistence is hypothesized to circumvent Eigen's paradox of the origin of life [7] —no large genome without enzymes, and no enzymes without a large genome— by assuming that, initially, each short template coded for a small piece of a modular enzyme (see, *e.g.*, [13]). The (haploid) individuals are viewed as self-replicating molecules or templates and the groups as protocells or vesicles that are themselves capable of reproduction. This is then a prototypical multilevel selection problem [14] which has been analyzed chiefly through Monte Carlo simulations and numerical iteration of recursion equations. The two dynamics that govern the competition between templates and between protocells are coupled because the reproduction rate of the protocells depends on their template composition.

The evolutionary processes we consider here are individual selection and group selection. In addition, since we assume that the population of templates inside each protocell is large but finite, random genetics drift plays an important role too, especially in the hindering of template coexistence. The population of protocells, however, is assumed infinite. We consider two distinct types of templates only and use the so-called diffusion approximation to obtain the partial differential equation that determines the proportion of protocells with a given composition of templates at a given time [15]. Analysis of the (singular) steady-state solution reveals a continuous phase transition separating the coexistence regime, in which a fraction of protocells exhibit a mixed composition of templates, from the segregation regime, in which there is no coexistence of template types inside

[(a)] On leave of absence from: Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica, Università dell'Aquila - Aquila, Italy, EU.

a protocell. More pointedly, we show that the order parameter vanishes linearly with the distance to the critical line, and derive an analytical expression for the critical steady-state solution.

The model. – The population is divided into an infinite number of protocells each of which carrying exactly N templates. There are two types of templates —type 1 and type 2— which differ only by their replication efficiencies: type 1 has a selective disadvantage $s>0$ relative to type 2. We denote by $x\in[0,1]$ the frequency of type-1 templates within a protocell (the frequency of type-2 templates is then $1-x$), and by $\phi(x,t)$ the probability density of protocells with a fraction x of type-1 templates. Here we assume that N is sufficiently large so that x can be treated as a continuous variable. Hence $\phi(x,t)\,\Delta x$ yields the proportion of protocells carrying type-1 templates with frequency lying in the range $(x,x+\Delta x)$. There is no mutation between the template types and there is no migration of templates between protocells.

Within each protocell, template reproduction follows the rules of the standard Wright-Fisher model [16]. In particular, we assume that the individual selection coefficients are $1-s$ and 1 for type-1 and type-2 templates, respectively. So, if the frequency of type-1 templates within a given protocell is x before reproduction, then that frequency becomes $\tilde{x}=x(1-s)/(1-sx)\approx x-sx(1-x)$ after reproduction. We assume, as usual, that s is on the order of $1/N$. This deterministic process is followed by the random sampling of N templates: the probability that the given protocell carries exactly $i=0,\ldots,N$ templates of type 1 is given by the binomial

$$\binom{N}{i}\tilde{x}^i\,(1-\tilde{x})^{N-i}\,, \tag{1}$$

so that the frequency of type-1 templates after deterministic selection and random sampling is $x'=i/N$. The way this procedure changes the probability density $\phi(x,t)$ is derived using the diffusion approximation of population genetics [16], which consists essentially of the calculation of the first two jump moments $\langle(x'-x)^n\rangle$ to obtain the drift ($n=1$) and the diffusion ($n=2$) terms of a Fokker-Planck equation (see eq. (4)).

The competition between protocells is taken into account as follows. Denoting by $c(x)$ the selection coefficient of a protocell with a fraction x of type-1 templates we have

$$\phi(x,t+\Delta t)=[\phi(x,t)+c(x)\,\phi(x,t)\,\Delta t]\,\zeta, \tag{2}$$

where ζ is such that $\int_0^1 \mathrm{d}x\phi(x,t+\Delta t)=1$, *i.e.*, $\zeta=1/[1+\bar{c}(t)\,\Delta t]$ with

$$\bar{c}(t)=\int_0^1 c(x)\,\phi(x,t)\,\mathrm{d}x. \tag{3}$$

Finally, taking the limit $\Delta t\to 0$ we obtain the change in the fraction of protocells due to intercell selection, $\Delta\phi=[c(x)-\bar{c}(t)]\,\phi(x,t)\,\Delta t$.

Combining the changes in ϕ due to the Wright-Fisher process and the intercell selection results in the equation [15]

$$\frac{\partial}{\partial t}\phi=\frac{1}{2}\frac{\partial^2}{\partial x^2}\,[b(x)\phi]-\frac{\partial}{\partial x}\,[a(x)\phi]+[c(x)-\bar{c}(t)]\,\phi, \tag{4}$$

where $b(x)=x(1-x)/N$, $a(x)=-sx(1-x)$, and $\bar{c}(t)$ is given by eq. (3). Equation (4) was first derived by Kimura aiming at studying the efficiency of group selection on the evolution and maintenance of an altruistic character [15]. In our notation, the altruists are the type-1 templates, which have a selective disadvantage relative to type 2 or nonaltruistic templates. In the altruism context, the intercell selection coefficient or group selection pressure must be an increasing function of x, so that the higher the frequency of altruists in a group, the greater is the selective advantage of the group. In particular, Kimura has chosen the linear dependence $c(x)\propto x$ and included the effects of mutation and migration in his analysis [15]. It should be noted, however, that the introduction of mutation and migration, which affects only the drift term $a(x)$ in eq. (4), actually greatly simplifies the analysis since these mixing processes guarantee the existence of a regular equilibrium distribution [17].

Here we consider the coexistence problem instead, which is considerably more taxing to group selection than the altruistic version summarized above. In fact, despite their handicap at the individual level, the altruistic templates have a nonzero probability of being fixed in small groups solely through the effect of random drift, whereas this very effect is a major pressure against the coexistence between different templates [11]. To favor coexistence we choose the intercell selection coefficient

$$c(x)=cx(1-x)\,, \tag{5}$$

which is maximum for well-balanced protocells at which $x=1/2$. Here c is a parameter on the order of $1/N$ that measures the intensity of the group selection pressure towards coexistence. Prescription (5) is built around the so-called metabolic model of template cooperation which assumes that the presence of the two functional template types in the same protocell is needed to assemble a nonspecific replicase which, in turn, plays an essential role in the template replication process [9–12]. Since the hookup of the replicase requires products from the two template types, its production rate is proportional to the concentration of the rare type, hence the requirement that $c(x)$ is maximized by well-balanced protocells. In addition, if a protocell lacks any template type it is considered unviable, *i.e.*, $c(0)=c(1)=0$. We note that in the metabolic model of cooperation each template type contributes indirectly to the replication of the other type through the catalytic action of the nonspecific replicase.

Analysis of the steady state. – At the steady state $\partial\phi/\partial t = 0$, eq. (4) reduces to

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left[x\left(1-x\right)\phi\right] + S\frac{\mathrm{d}}{\mathrm{d}x}\left[x\left(1-x\right)\phi\right] + \left[Cx\left(1-x\right) - \bar{C}\right]\phi = 0, \tag{6}$$

where $S = 2Ns$ and $C = 2Nc$ are now parameters that can take on arbitrary positive values, and

$$\bar{C} = C\int_0^1 x\left(1-x\right)\phi\left(x\right)\mathrm{d}x. \tag{7}$$

The solution ϕ has to be found in the interval $[0,1]$ and since the extremes of this interval are absorbing barriers it can be written in the general form

$$\phi\left(x\right) = A_0\,\delta\left(x\right) + A_1\,\delta\left(1-x\right) + B\eta\left(x\right), \tag{8}$$

where A_0, A_1 and B are positive weights. Here $\eta\left(x\right)$ is a regular function that satisfies the second-order differential equation (6) in the open interval $(0,1)$ and can be eventually continued in the interval $[0,1]$ by defining $\eta\left(0\right) = \lim_{x\to 0}\,\eta\left(x\right)$ and $\eta\left(1\right) = \lim_{x\to 1}\,\eta\left(x\right)$. In addition, imposing the normalization $\int_0^1 \eta\left(x\right)\mathrm{d}x = 1$ we have

$$A_0 + A_1 + B = 1 \tag{9}$$

and

$$\bar{C} = BC\int_0^1 x\left(1-x\right)\eta\left(x\right)\mathrm{d}x. \tag{10}$$

We note that eq. (8) implies that, in the general case, the population is composed of three types of protocells: i) homogeneous protocells carrying type-1 templates only; ii) homogeneous protocells carrying type-2 templates only; and iii) heterogeneous protocells carrying any arbitrary mixture of the two templates. The proportion of each type in the infinite protocell population is A_0, A_1 and B, respectively.

Next, we derive two most useful relations between the values of $\eta\left(x\right)$ at the extremes $x = 0$ and $x = 1$ and the weights that appear in eq. (8). Integrating eq. (6) over the interval $[-\epsilon,\epsilon]$ and neglecting terms of order of ϵ and higher yield $A_0 = B\eta\left(0\right)/\bar{C}$ or, equivalently,

$$A_0 = \frac{\eta\left(0\right)}{C\int_0^1 x\left(1-x\right)\eta\left(x\right)\mathrm{d}x}, \tag{11}$$

where we have used $\frac{\mathrm{d}}{\mathrm{d}x}\left[x(1-x)\eta(x)\right] = x(1-x)\frac{\mathrm{d}}{\mathrm{d}x}\eta(x) + (1-2x)\eta(x)$, $\eta(\epsilon) = \eta(0) + o(\epsilon)$ and $\eta(-\epsilon) = 0$. Here $o(\epsilon)$ contains all terms of order ϵ or higher. Similarly, integration of eq. (6) over the interval $[1-\epsilon, 1+\epsilon]$ yields

$$A_1 = \frac{\eta\left(1\right)}{C\int_0^1 x\left(1-x\right)\eta\left(x\right)\mathrm{d}x}. \tag{12}$$

We note that $B = 1 - A_0 - A_1$ can be inserted in eq. (10) so that the regular solution η can be obtained autonomously (see eq. (15) for a similar operation).

The numerical procedure to find the regular solution $\eta\left(x\right)$ of eq. (6) is greatly facilitated if we define the auxiliary function $y\left(x\right)$ as

$$\eta\left(x\right) = R\,\exp\left(-Sx/2\right)\,y\left(x\right), \tag{13}$$

where $R = 1/\int_0^1 \mathrm{d}x\exp\left(-Sx/2\right)y\left(x\right)$ is introduced to guarantee the normalization of η, leaving us free to impose the normalization $\int_0^1 \mathrm{d}x\,y(x) = 1$ on the auxiliary function y. It is easy to verify from eq. (6) that in the open interval $(0,1)$ we have

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left[x\left(1-x\right)y\right] + \Gamma x\left(1-x\right)y = \bar{C}y \tag{14}$$

where $\Gamma = C - S^2/4$ and $\bar{C}$ is obtained by integration of eq. (14) in the interval $(0,1)$ as

$$\bar{C} = \Gamma\int_0^1 x\left(1-x\right)y\left(x\right)\mathrm{d}x - y\left(0\right) - y\left(1\right) \tag{15}$$

where $y(0) = \lim_{x\to 0}y(x)$ and $y(1) = \lim_{x\to 1}y(x)$. Since the differential equation (14) is symmetrical around $x = 1/2$, we assume that the solution $y\left(x\right)$ is symmetrical too, so that $y(0) = y(1)$.

Numerical solution. – At this stage our mathematical problem is ready for a numerical approach. We are left with a Sturm-Liouville problem, eq. (14), without boundary conditions which can be solved by requiring the regularity of $y\left(x\right)$ in $[0,1]$ only [18]. Of course, this requirement is satisfied provided that $\bar{C}$ is the eigenvalue of the Sturm-Liouville problem, which depends solely on the value of Γ. In practice, we determine $\bar{C}$ by propagating the solution using the Runge-Kutta algorithm from $x = 0$ to $x = 1/2$, from $x = 1$ to $x = 1/2$, and then requiring that the two solutions join smoothly at $x = 1/2$. We use an arbitrary value for $y(0) = y(1)$ to begin the Runge-Kutta iterations, and the first derivatives $y'(0) = -y'(1) = \left(1 + \bar{C}/2\right)y(0)$ as given by eq. (14). The arbitrariness of the initial condition has no effect on the resulting eigenvalue $\bar{C}$ and it is mitigated by imposing the normalization condition on $y(x)$. In fact, since the coefficients A_0 and A_1 given by eqs. (11) and (12), as well as $B = 1 - A_0 - A_1$, depend on the ratio of terms involving $\eta\left(x\right)$ (and so $y\left(x\right)$) our arbitrary choice of $y(0)$ is inconsequential. Finally, we note that eq. (15) does not provide any additional information about y or $\bar{C}$ —it is derived from eq. (14)— and so it is not used in our numerical procedure.

The dependence of the eigenvalue $\bar{C}$ on Γ exhibited in fig. 1 reveals the existence of a critical value Γ_c below which $\bar{C} = 0$. This critical point separates two distinct steady-state regimes regarding the possibility of coexistence of the two types of templates within the same protocell, and so $\bar{C}$ can be seen as the order parameter of our problem. We note that $\bar{C}/\Gamma$ (the coefficient of the linear term in the expansion of $\bar{C}$ in powers of Γ) varies very slowly with increasing Γ, being confined to the interval $[0.205, 0.25]$. In fact, since in the limit of large Γ

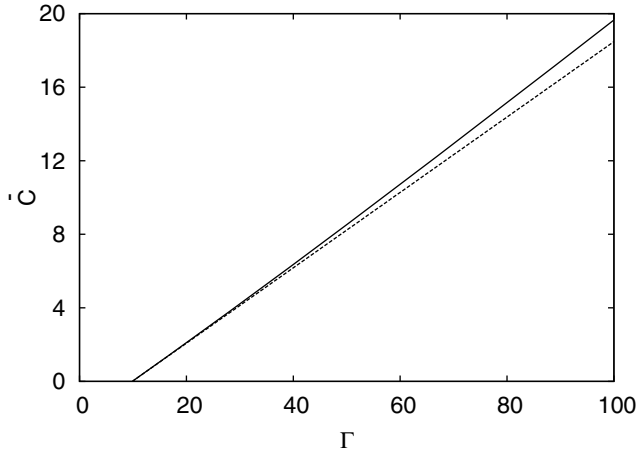


Fig. 1: Eigenvalue $\bar{C}$ as a function of $\Gamma = C - S^2/4$. For $\Gamma \leqslant \Gamma_c = \pi^2$ we find $\bar{C} = 0$, which characterizes the segregation regime. The dashed line is $\bar{C} = \alpha(\Gamma - \Gamma_c)$ with α given by eq. (28).

we have $\Gamma \approx C \gg 1$ the group selection pressure favoring coexistence becomes the dominant force leading to an ideal coexistence scenario, *i.e.*, $\phi = \delta(x - 1/2)$, for which $\bar{C}/\Gamma \approx 1/4$ (see eq. (7)).

Before we carry on the characterization of the two steady-state regimes separated by Γ_c, let us show how the value of this critical parameter can be calculated.

The transition line. – Setting $\bar{C} = 0$ and defining $\psi(x) = x(1-x)\,y(x)$, eq. (14) reduces to the harmonic-oscillator equation $\mathrm{d}^2\psi/\mathrm{d}x^2 + \Gamma_c\psi = 0$. The solution $\psi = \kappa_c \cos(\Gamma_c^{1/2}x) + \kappa_s \sin(\Gamma_c^{1/2}x)$ must vanish at $x = 0$ and $x = 1$, otherwise $y(x)$ would not be normalizable. This implies that $\kappa_c = 0$, and $\Gamma_c = \pi^2$ or, equivalently,

$$C_c = \pi^2 + \frac{S^2}{4}. \tag{16}$$

Hence, at the critical point we have $y_c(x) = \kappa_s \sin(\pi x)/[x(1-x)]$ where the normalization factor is $\kappa_s = 1/\int_0^1 \mathrm{d}x \sin(\pi x)/[x(1-x)] \approx 0.270$. Finally, returning to the original regular probability density we write

$$\eta_c(x) = R_c\kappa_s \exp(-Sx/2)\,\frac{\sin(\pi x)}{x(1-x)}, \tag{17}$$

where R_c is the normalization constant that appear in eq. (13) evaluated at $\Gamma = \Gamma_c$.

As a consistency check we will show now that $\bar{C} = 0$ at $C = C_c$. In fact, eq. (17) yields $\eta_c(0) = R_c\kappa_s\pi$, $\eta_c(1) = R_c\kappa_s\pi\exp(-S/2)$, and

$$\int_0^1 x(1-x)\,\eta_c(x)\,\mathrm{d}x = \frac{R_c\kappa_s\pi}{C_c}\left[1 + \exp(-S/2)\right]. \tag{18}$$

Inserting these results in eqs. (11) and (12) yields

$$A_0^c = \frac{1}{1 + \exp(-S/2)} \tag{19}$$

and

$$A_1^c = \frac{\exp(-S/2)}{1 + \exp(-S/2)}. \tag{20}$$

Since $A_0^c + A_1^c = 1$ it follows from eqs. (9) and (10) that $B^c = 0$ and $\bar{C} = 0$.

Characterization of the steady-state regimes. – The first regime is associated to the parameter region $C \leqslant C_c$, and is characterized by $\bar{C} = 0$. According to eq. (7), its protocell distribution ϕ is a sum of Dirac deltas centered at $x = 0$ and $x = 1$, *i.e.*, $B = 0$ in eq. (8). Since the two types of templates do not coexist within the same protocell we refer to this regime as the segregation regime. This is a nonergodic regime where the weights A_1 and $A_0 = 1 - A_1$ depend on the initial conditions, *i.e.*, on the distribution $\phi(x, 0)$. In the following we calculate this dependence explicitly in the case $S = 0$.

Conforming to the previous rescalings $C = 2Nc$ and $S = 2Ns$, we begin by replacing t by $2Nt$ in eq. (4). In addition, to lighten the notation we introduce the abbreviation $\langle f(x)\rangle_t = \int_0^1 f(x)\,\phi(x,t)\,\mathrm{d}x$ for the expected value of a regular function $f(x)$ at time t. Hence

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle f(x)\rangle_t = \left\langle x(1-x)\,\frac{\partial^2 f(x)}{\partial x^2}\right\rangle_t - \bar{C}(t)\langle f(x)\rangle_t + C\langle x(1-x)\,f(x)\rangle_t \tag{21}$$

with $\bar{C}(t) = C\langle x(1-x)\rangle_t$. We recall that $\lim_{t\to\infty}\bar{C}(t) = \bar{C} = 0$ in the segregation regime. The choice $f(x) = \sin\left(C^{1/2}x + \theta\right)$, where θ is an arbitrary constant, allows us to rewrite eq. (21) as

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\sin(C^{1/2}x + \theta)\rangle_t = -\bar{C}(t)\langle\sin(C^{1/2}x + \theta)\rangle_t, \tag{22}$$

whose formal solution is

$$\frac{\left\langle\sin\left(C^{1/2}x + \theta\right)\right\rangle_t}{\left\langle\sin\left(C^{1/2}x + \theta\right)\right\rangle_0} = \exp\left[-\int_0^t \bar{C}(\tau)\,\mathrm{d}\tau\right]. \tag{23}$$

Since the right-hand side of this equation does not depend on θ, neither does the left-hand side so we can write

$$\frac{\left\langle\sin\left(C^{1/2}x\right)\right\rangle_t}{\left\langle\sin\left(C^{1/2}x\right)\right\rangle_0} = \frac{\left\langle\cos\left[C^{1/2}(x - 1/2)\right]\right\rangle_t}{\left\langle\cos\left[C^{1/2}(x - 1/2)\right]\right\rangle_0} \tag{24}$$

for the choices $\theta = 0$ and $\theta = \left(\pi - C^{1/2}\right)/2$. This identity holds for any t and therefore it holds also in the limit $t \to \infty$ at which $\lim_{t\to\infty}\phi(x,t) = \phi(x) = A_0\delta(x) + A_1\delta(x-1)$. The evaluation of the averages in this limit and the use of $A_0 = 1 - A_1$ yield

$$A_1 = \frac{\cos\left(C^{1/2}/2\right)\int_0^1 \sin\left(C^{1/2}x\right)\phi(x,0)\,\mathrm{d}x}{\sin\left(C^{1/2}\right)\int_0^1 \cos\left[C^{1/2}(x - 1/2)\right]\phi(x,0)\,\mathrm{d}x}, \tag{25}$$

where we have reverted to the original integral notation to emphasize the dependence on $\phi(x, 0)$. The classic result $A_1 = \int_0^1 x\phi(x,0)\,\mathrm{d}x$ is recovered for $C \to 0$ [16] as well

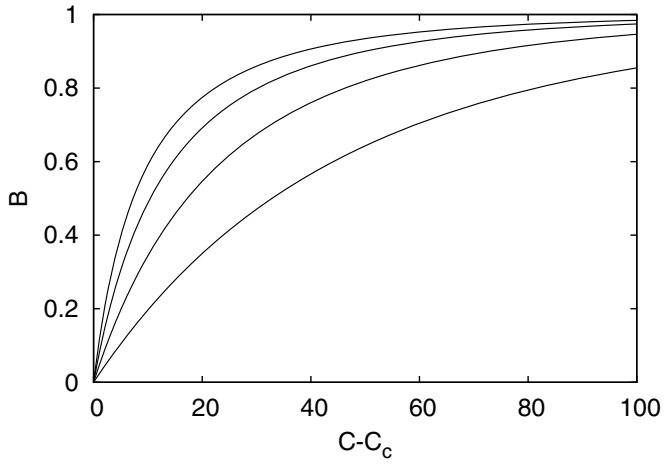


Fig. 2: Fraction of protocells carrying the two template types as a function of the distance to the critical point $C - C_c = \Gamma - \Gamma_c$ for (top to bottom) $S = 0, 5, 10$ and 20.

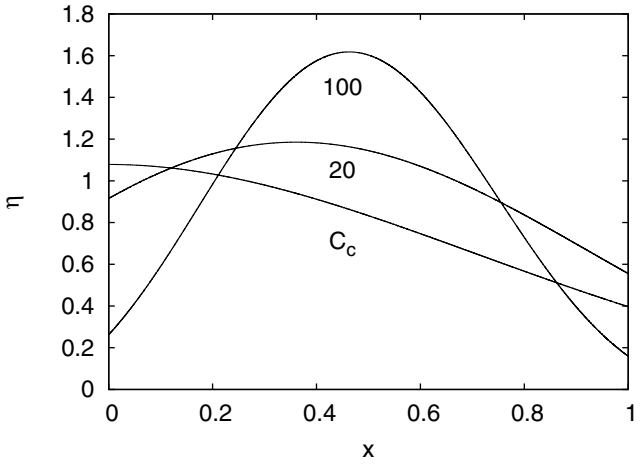


Fig. 3: Regular normalized solution $\eta(x)$ for $S = 1$ and $C = C_c \approx 10.12$, 20 and 100 as indicated in the figure.

as $A_1 = A_1^c = 1/2$ for $C \to \pi^2$ (see eq. (20)). Hence A_1 is continuous at the critical point $C = C_c$. Although this approach allows the calculation of the weight A_1 for $S > 0$ as well, the derivation is too lengthy and will be omitted here.

The second steady-state regime is characterized by $\bar{C} > 0$ and so admits the existence of protocells in which the two types of templates coexist. The quantity of interest here is the weight B (see eq. (8)) which gives the fraction of protocells carrying the two template types, regardless of their frequencies inside the protocells. In fig. 2 we present the dependence of B on the distance to the critical point $\Gamma - \Gamma_c = C - C_c$ for different values of the (rescaled) individual selection coefficient S. Note that although $\bar{C}$ depends on Γ only, the weights A_0, A_1 and B associated to the different protocell types exhibit an explicit dependence on S as well, which comes through the function η (see eq. (13)). In particular, for a fixed group selection pressure, increase of the selective advantage of type-2 templates reduces the fraction B of protocells that exhibit coexistence between templates, as expected. More pointedly, we will show that, close to the critical point, B decreases with $1/S$ for increasing S.

In fig. 3 we show the regular part of the density of probability of protocells $\eta(x)$ for different group selection intensities. The protocells are unbalanced in this figure because of the choice $S = 1$ which confers a selective advantage to type-2 templates in the competition at the individual level that goes on inside the protocells. However, as C increases this unbalance diminishes and eventually the population is completely dominated by well-balanced protocells so that $\eta(x) \to \delta(x - 1/2)$.

Next we derive explicit analytical expressions for the behavior of the order parameter $\bar{C}$ as well as for the coexistence probability B near the critical point Γ_c. Multiplying eq. (14) by $\exp(-\lambda x)$ and integrating in the interval $(0, 1)$ yield

$$\bar{C}/Q = \left(\Gamma + \lambda^2\right) \int_0^1 \exp(-\lambda x)\, x\,(1-x)\, y(x)\, \mathrm{d}x - y(0)\left[1 + \exp(-\lambda)\right] \tag{26}$$

with $Q = 1/\int_0^1 \mathrm{d}x \exp(-\lambda x)\, y(x)$. This equation reduces to eq. (15) when $\lambda = 0$, whereas, for $\lambda = S/2$, it coincides with the relation that is obtained by combining eqs. (9), (10), (11), (12) and (13) and taking into account that $\Gamma = C - S^2/4$.

Taking $\lambda = i\pi$ and recalling that $\Gamma_c = \pi^2$, we obtain

$$\bar{C} = (\Gamma - \Gamma_c)\, \frac{\int_0^1 x\,(1-x) \sin(\pi x)\, y(x)\, \mathrm{d}x}{\int_0^1 \sin(\pi x)\, y(x)\, \mathrm{d}x}, \tag{27}$$

where we have used the fact that $y(x)$ is symmetric about $x = 1/2$. For Γ close to Γ_c we can replace y by y_c and write $\bar{C} \approx \alpha(\Gamma - \Gamma_c)$, where

$$\alpha = \left[2 \int_0^1 \frac{\sin^2(\pi x)}{x\,(1-x)} \mathrm{d}x\right]^{-1} \approx 0.205. \tag{28}$$

Now it is easy to obtain the behavior of the coexistence probability B near the critical point. In fact, once the behavior of $\bar{C}$ is known, use of eqs. (10) and (18) allows us to write $B \approx \beta(\Gamma - \Gamma_c)$, where

$$\beta = \frac{\alpha}{\pi\left[1 + \exp(-S/2)\right]} \int_0^1 \exp(-Sx/2)\, \frac{\sin(\pi x)}{x\,(1-x)} \mathrm{d}x. \tag{29}$$

For $S = 0$ we have $\beta = \alpha/2\pi\kappa_s \approx 0.121$, whereas for $S \gg 1$ we find $\beta \approx 2\alpha/S$.

Conclusion. – Group selection and, more generally, structured population arguments aiming at explaining altruistic behavior and eusociality in nature have been a source of controversy since they were first proposed in

the 1960s [19] (see [20] and accompanying refutations for a recent clash). However, use of group selection ideas to explain the coexistence of selfish individuals is a much less loaded issue, at least within the prebiotic evolution context, since there seems to exist a consensus that the compartmentalization of templates was a necessary evolutionary step towards the modern cell [21]. Regardless of the relevance and controversies surrounding the role of group selection in nature, the challenging mathematical models used to describe the resulting multilevel selection problem have been viewed as an attraction on their own [15,22,23]. Here we follow this tradition and offer a numerical and analytical solution to a nontrivial group selection model.

We build on the work of Kimura [15] and consider a group selection pressure towards the coexistence of two types of templates with distinct replication rates. We find an explicit expression for the critical intensity of group selection needed to balance the segregation effects of genetic drift and individual selection, thus guaranteeing the existence of protocells carrying the two template types (see eq. (16)). In addition, we derive analytical expressions for the steady-state distribution at the critical point and show that the transition between the regimes of coexistence and segregation is continuous with the order parameter vanishing linearly with the distance to the critical point. The model is nontrivial because it preserves the main effect of finite populations —genetic drift— which is revealed when the values of the rescaled parameters $C = 2Nc$ and $S = 2Ns$ decrease towards zero. Since $C_c \to \pi^2$ in this case, template coexistence is ruled out for small protocells due to the segregating effect of genetic drift.

We note that the actual value of the population size N has little relevance in the diffusion approximation framework, provided it is large enough to allow treating the frequency x as a continuous variable. The relevant quantities that appear in the equations are always expressed as products between N and the original parameters (s and c in our case), and only them are accessible to experiments involving small population samples (see Chapt. 13 of [24]). Nevertheless, we can speculate on the value of N using the recent estimate that a human cell infected with the retrovirus HIV-1 may produce more than 10^4 new viral particles over its short life span [25]. This figure may then be viewed as a rough estimate of the carrying capacity of the protocells.

From a mathematical perspective, our work departs from previous population genetic studies using the diffusion approximation due to the absence of the mixing and regularizing processes of mutation and migration [16]. Inclusion of these processes would guarantee the existence of a regular equilibrium distribution for the Fokker-Planck-like equation (4). Use of prescription (8), however, allows us to single out the singularities at the absorbing barriers $x = 0$ and $x = 1$, whereas the relations (11) and (12) provide a link between the regular part of the equilibrium distribution and the weights of the singular parts. This solution strategy can be applied to a variety of problems characterized by the superposition of absorbing and extended steady states.

* * *

The authors are grateful to Luca Peliti for having spurred their collaboration and for having highlighted the issues yet to be clarified about the role of altruism in evolutionary processes. JFF was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) while MS was partially supported by PRIN 2009 protocollo n. 2009TA2595.02.

REFERENCES

[1] Hutchinson G. E., *Am. Nat.*, **95** (1961) 137.
[2] Hardin G., *Science*, **162** (1968) 1243.
[3] Eigen M., *Naturwissenschaften*, **58** (1971) 465.
[4] Károlyi G., Péntek A., Scheuring I., Tél T. and Toroczkai Z., *Proc. Natl. Acad. Sci. U.S.A.*, **97** (2000) 13661.
[5] Hamilton W. D., *J. Theor. Biol.*, **7** (1964) 1.
[6] Trivers R. L., *Q. Rev. Biol.*, **46** (1971) 35.
[7] Eigen M. and Schuster P., *Naturwissenschaften.*, **64** (1977) 541.
[8] Wilson D. S., *The Evolution of Populations and Communities* (Benjamin Cummings, Menlo Park) 1980.
[9] Niesert U., Harnasch D. and Bresch C., *J. Mol. Evol.*, **17** (1981) 348.
[10] Szathmáry E. and Demeter L., *J. Theor. Biol.*, **128** (1987) 463.
[11] Fontanari J. F., Santos M. and Szathmáry E., *J. Theor. Biol.*, **239** (2006) 247.
[12] Silvestre D. A. M. M. and Fontanari J. F., *Phys. Rev. E*, **75** (2007) 051909.
[13] Manrubia S. C. and Briones C., *RNA*, **13** (2007) 97.
[14] Traulsen A., Sengupta A. M. and Nowak M. A., *J. Theor. Biol.*, **235** (393) 2005.
[15] Kimura M., *Proc. Natl. Acad. Sci. U.S.A.*, **80** (1983) 6317.
[16] Crow J. F. and Kimura M., *An Introduction to Population Genetics Theory* (Harper & Row, New York) 1970.
[17] Wright S., *Genetics*, **16** (1931) 97.
[18] Chalub F. A. C. C. and Souza M. O., *Commun. Math. Sci.*, **7** (2009) 489.
[19] Williams G. C., *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Princeton University Press, Princeton) 1966.
[20] Nowak M. A., Tarnita C. E. and Wilson E. O., *Nature*, **466** (2010) 1057.
[21] Eigen M., Gardiner W. C. jr. and Schuster P., *J. Theor. Biol.*, **85** (1980) 407.
[22] Aoki K., *Evolution*, **36** (1982) 832.
[23] Donato R., Peliti L. and Serva M., *Theor. Biosci.*, **116** (1997) 309.
[24] Nei M., *Molecular Evolutionary Genetics* (Columbia University Press, New York) 1987.
[25] De Boer R. J., Ribeiro R. M. and Perelson A. S., *PLoS Comput. Biol.*, **6** (2010) e1000906.