A Biological Generator of Prime Numbers

Eric Goles[(1)], Oliver Schulz[(2)], and Mario Markus[(3)]

[(1)] *Center for Mathematical Modelling of Complex Systems, FCFM, University of Chile, Casilla 170-3, Santiago, Chile Systems, FCFM, University of Chile*
[(2),(3)] *Max-Planck-Institut fuer molekulare Physiologie, Postfach 102664, D-44026 Dortmund, Germany*

In the present work, we shall allow for the merging of two seemingly unrelated subjects, one being periodical insects and the other the theory of prime numbers. The fact that some species of cicadas (genus Magicicada) appear every 7, 13 or 17 years and that these periods are prime numbers has been regarded as a coincidence [1, 2]. Without intending to argue in favour or against this statement, we will show here that a simple evolutionary predator-prey model yields prime-periodic preys. Moreover, this result will be used as a number-theoretical tool, namely to generate large prime numbers. Furthermore, we will demonstrate how a spatio-temporal extension of the model renders spiral waves being reminiscent of those observed in excitable systems, host-parasitoid systems and prebiotic evolution.

Key words: prime numbers, prebiotic evolution, evolutionary predator-prey model, host-parasitoid system

PACS numbers: 87.23

1 Introduction

Scanning of the literature shows that the appearance of three species of the genus Magicicada with periods that are prime numbers has not yet been undersstood [1, 2, 3, 4, 5]. These cicadas spend most of their lives below the ground as nymphs, emerging and dying within a few weeks. A few attempts have been made to model this phenomenon: Hoppenstedt and Keller [6] presented a predator-prey model which - after a delicate adjustment of parameters - shows that synchronized periodical behaviour is possible for periods longer than 10 years; Bulmer [7] obtained similar results considering mainly the competition of nymphs below the ground. An essential point of the present work is the assumption that the cycle length is a prime number in order to optimally escape predators; this is because preys are subject to predators whose cycle lengths are divisors of the cycle of the preys. Lloyd and Dybas [8] hold this view by mentioning hypothetical parasitoids, attacking eggs or adults, that may have become extinct. Non-extinct examples of non-annual predators are: the cicada-hunting wasp Exeirus lateritius (dormancy 2 or 3 years) and the fungus Massospora cicadina, which also attacks adult cicadas and may be dormant for several decades.

2 Temporal model

In our model we consider a predator of period X interacting with a prey of period Y. We assign a momentary fitness $f_x(t)$ of the predator in a year t as follows: it is zero if it is not present, it is $+1$ if both predator and prey are present, and it is -1 if the predator is present but the prey is not. The momentary prey fitness $f_y(t)$ is defined analogously, but with opposite signs. The fitness F_x, resp. F_y, is defined by the sum over the $f_x(t)$, resp. $f_y(t)$, $t = 0, \ldots, XY$, divided by the number of predator, resp. prey, generations. (Note that this yields an average valid for $t \longrightarrow \infty$, since the process is peri-
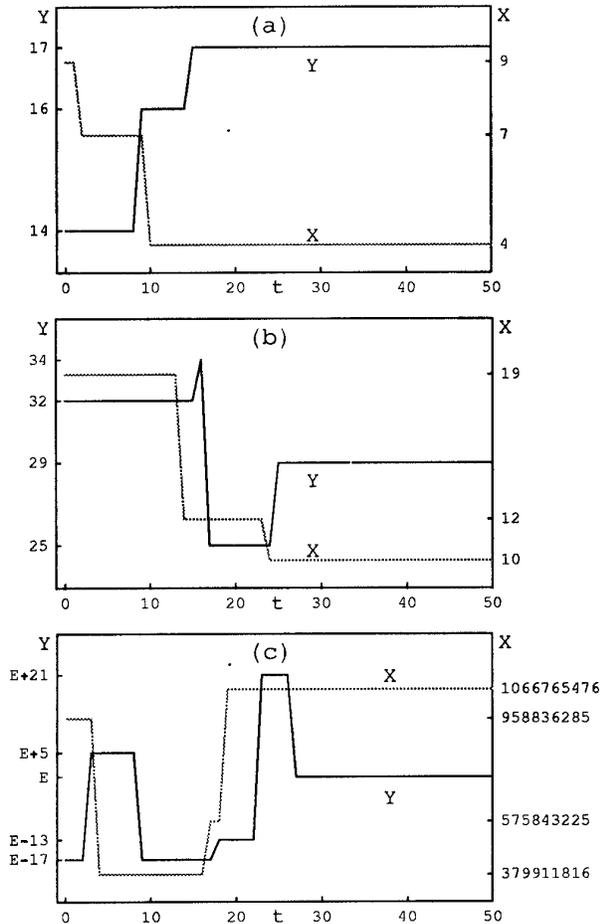
FIG. 1. Evolution to the prey-period $Y = 17$ (a), to $Y = 29$, and to $Y = E = 2147483647$ (c; prime number discovered by Euler) through mutation-selection-sequences (abscissa: number of time steps t; left ordinate: prey period Y; right ordinate: predator period X).

odic with period XY). We divide by the number of generations in order to avoid selection by virtue of the capability for frequent proliferation; we rather assume that each generation uses up metabolic resources – due to metamorphosis, mating and death – and these expenses should be minimized in the long run. We compare now a prey mutating to a cycle Y' with the resident prey (cycle length Y) at constant X. Analogously, we compare mutant cycles

X' with resident cycles X at constant Y. A mutant prey (resp. predator) substitutes the resident if and only if $F_{y'} > F_y$, resp. $F_{x'} > F_x$. Thus, in the case of fitness equality, the resident and not the mutant is selected. Here and in the rest of this work, we assume that all interacting populations are present at $t = 0$, and are thus synchronized.

Now we proceed to analyze the model described in the previous paragraph. Let $lcm(X,Y)$: least common multiple, $gcd(X,Y)$: greatest common divisor of X and Y. In XY years, the predator appears Y times, both predator and prey appear $XY/lcm(X,Y)$ times, thus predator without prey appears $Y - XY/lcm(X,Y)$ times. Considering that $gcd(X,Y) \cdot lcm(X,Y) = XY$, we thus obtain the predator fitness

$$F_x(X,Y) = 2\,gcd(X,Y)/Y - 1.$$

Analogously, one obtains the prey fitness

$$F_y(X,Y) = 1 - 2\,gcd(X,Y)/X.$$

We will now show the following: if we allow arbitrary mutations leading to periods within the ranges $2 \le X \le L/2 + 1$, $L/2 + 2 \le Y \le L$ (we call H the rectangle defined by these cycle-length limiting ranges on the X-Y-plane), then a random sequence of mutations will finally lock the prey into a stable prime period Y. More precisely, we will show that if Y is not a prime then there exists a sequence of mutations that will change Y, and if Y is prime then no mutation will change it. Let us first assume that $Y = Y_N$ is not a prime; $F_x(X, Y_N)$ has the maximum value $2\,gcd(X_M, Y_N)/Y_N - 1$ at the predator period $X_M = gd(Y_N)$ (gd: greatest divisor). Note that $1 < X_M < L/2 + 1 \Rightarrow (X_M, Y_N) \in H$. A sequence of random mutations keeping $Y = Y_N$ constant will eventually lead to X_M. However, (X_M, Y_N) is abandoned if mutations lead to $(X_M, Y_N \pm 1)$. In fact, $gcd(X_M, Y_N) = X_M$, implying that $F_y(X_M, Y_N) = -1$; $gcd(X_M, Y_N \pm 1)$ cannot be equal to X_M (the reason is: $(Y_N \pm 1)/X_M = Y_N/X_M \pm 1/X_M$, the first term being an integer, but the second not, so that X_M is not a divisor of $Y_N \pm 1$) and $gcd(X_M, Y_N \pm 1)$ can certainly not be larger than X_M; thus $gcd(X_M, Y_N \pm 1) < X_M$; this

implies that $F_y(X_M, Y_N \pm 1) > -1 = F_y(X_M, Y_N)$. So far, we have shown that there exists a sequence of mutations such that a prey with a non-prime cycle $Y = Y_N$ is extinguished. Assume now that $Y = Y_P$ is a prime; any $X = X_R$ such that $(X_R, Y_P) \in H$ satisfies the condition $1 \le X_R \le Y_P - 1$ and is thus relatively prime to Y_P; therefore $gcd(X_R, Y_P) = 1$, so that starting from $(X_R, Y_P) \in H$, there exist no predator mutants that are fitter than a resident. On the other hand, $gcd(X_R, Y') \ge 1$, where Y' is a prey mutant, as compared to $gcd(X_R, Y_P) = 1$, so that $F_y(X_R, Y') \le F_y(X_R, Y_P)$, i.e. no prey mutant is fitter than a resident. In conclusion, if H contains a prime prey cycle length, any initial random choice of $(X, Y) \in H$ will lead and lock to a prime Y after a sufficiently large number of mutations. This locking to primes is illustrated in Fig. 1 for a case of biological relevance (Fig. 1a; locking into $Y = 17$) and for the two cases. (Fig. 1b and 1c). In particular, locking of Y in Fig. 1c occurs into the Euler-prime E.

At this point we would like to comment on our restriction to the above rectangle H. We imposed this restriction because the points (jY_P, Y_P), where Y_P is prime and $j = 1, 2, 3, \ldots$, are unstable with respect to prey mutations. This can be shown easily: $gcd(jY_P, Y_P) = Y_P$, while $gcd(jY_P, Y_P - k)$ with $k = 1, 2, 3, \ldots$ cannot be larger than $Y_P - k$; thus $F_y(jY_P, Y_P) < F_y(jY_P, Y_P - k)$.

This means that convergence to a prey with period Y_P is not possible if mutations to the points (jY_P, Y_P) are permitted. In order to discard these points, one could restrict the system to points above the diagonal X=Y. This, however, is biologically not plausible, since it would mean that the limits for prey cycle lengths depend on predator cycles lengths, and viceversa. Such dependencies are avoided by restricting mutations to the rectangle H. In addition, H fulfills the requirement that it contains $X_M = gd(Y_N)$ (see above). Note that we allow arbitrary mutations within H; this implies that if L is chosen sufficiently large, an evolution with initial conditions with low values for X and Y (such as in Fig. 1a) may be "locked" in a large prime prey cycle (such as in Fig. 1c).
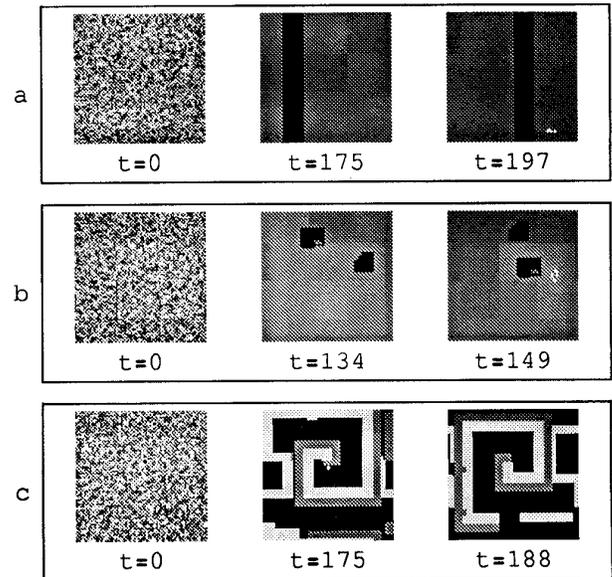
3   Spatio-temporal model

FIG. 2. Spatiotemporal dynamics of the cellular automaton starting from a random distribution and evolving to gliders (a,b) and to a spiral surrounded by pulsating smaller domains (c). At $t = 0$, one predator (with cycle X) and one prey (with cycle Y) are placed in each cell by random choice within $\sqrt{E} < X < E - 2$, $E - 8 < Y < E + 8$, where E is the Euler-prime (see Fig. 1c). Number of cells: 64×64. Only the prey is displayed. Different gray shadings correspond to different Y, black corresponding to $Y = E$.

In what follows, we consider competition between neighbouring residents in a spatially extended system, instead of competition between mutants and residents. Fig. 2 shows results obtained with a cellular automaton (CA). CA have been shown to be useful tools for simulations of natural phenomena in space and time [18, 19, 20, 21]. The CA here evolves in a two-dimensional habitat, as follows. In each time step, the predator and the prey of each cell are replaced by the fittest among the neighbours. The neighbourhood is defined by the cell itself and the 8 cells around it. The momentary fitness $f_x(t)$ of a predator in the time step t is computed here as

follows: $f_x(t) = 0$ if the predator does not appear in that time step; if the predator appears and the number ν of cells in the neighbourhood occupied by prey is not zero ($1 \leq \nu \leq 9$), then $f_x(t) = \nu$; $f_x(t) = -p$ if the predator appears and $\nu = 0$. p is an integer describing a "punishment" for a predator that appears but finds no prey at all. The momentary fitness $f_y(t)$ of the prey is computed analogously, but with opposite signs. The fitness F_x, resp. F_y, of a predator, resp. prey, are given by the sum of the $f_x(t)$, resp $f_y(t)$, over all t, t ranging from 1 to the product of all interacting cycle lengths; this sum is then divided by the number of generations of the predator, resp. the prey. The predator, resp. the prey, in a cell are replaced in the next time step by the predator, resp. prey, having the largest fitness F_x, resp. F_y, within the neighbourhood. We use cyclic boundary conditions.

The CA just described yields a large diversity of coexisting attractors, depending solely on the choice of the initial, randomly distributed cycle lengths. As in the "Prisoner's dilemma" CA in Ref. 17, the fate of each cell depends here on 25 neighbours; in fact, the fitness of each cell is evaluated by the encounters among the inhabitants of this cell and of its 8 neighbours, but after each time step the inhabitants of a cell are replaced by those having the largest fitness, considering the fitness of that cell and that of each of its 8 neighbours. This contrasts with Conway's "Game of Life" [9], where 9 cells specify a cell's fate. After a sufficient number of time steps, we obtain homogeneity, travelling waves (including "gliders" such as those shown in Fig, 2a and 2b), spiral waves (Fig. 2c) and periodical, complicatedly pulsating areas, such as those surrounding the spiral in Fig. 2b. We found a predominant appearance of prime-periodic preys, either in the waves or in the background.

The wave in Fig. 2a can be better understood by considering a CA with only one spatial dimension. (Note that for graphical simplicity, only the prey is displayed in Fig. 2.). We call the cycles in the background X_B, Y_B and those in the wave X_W, Y_W . A 1D-wave is composed of five zones defined on the cells denoted by $i = 1, 2, \ldots, n, \ldots n+m, \ldots$, where $m > 2$ is the width of the wave: i) X_B, Y_B at $i = 1, 2, \ldots, n$;   ii) X_B, Y_W at $i = n+1$;   iii) X_W, Y_W at $i = n+2, \ldots, n+m$;   iv) X_W, Y_B at $i = n+m+1$;   v) X_B, Y_B at $i = n+m+2, \ldots$. Note that the predator-wave is displaced one cell to the right relatively to the prey-wave; this determines the moving direction to the right, as we will explain now. Considering that predator-prey interactions occur here only with the two immediate neighbours of each cell, the predator at $i = n+m+1$ (resp. at $i = n+1$) can feed on two types of prey and thus has a larger fitness than the predator at $i = n+m+2$ (resp. at $i = n+2$), which can only feed on one type of prey; therefore, the predator-wave will move one cell to the right in the next time step. The prey at $i = n+m+1$ (resp. the prey at $i = n+1$) can be eaten by two types of predators and thus has a lower fitness than the prey at $i = n+m$ (resp. at $i = n$), which can only be eaten by one type of predator; therefore also the prey-wave will move one cell to the right. In two dimensions the mechanism is more complicated (especially for gliders moving diagonally, as in Fig. 2b) but they can be understood by the same type of reasoning steps. It is remarkable that one-dimensional waves and spiral waves here show a similar behaviour to waves in prebiotic evolutionary automata [11], in host-parasitoid dynamics [10], as well as in excitable media [12, 13, 14, 15], such as chemical reactions, heart muscle and epidemics, in spite of the mechanistic differences.

4   Statistical analysis

We now determine the probabilities $P(Y)$ that different prime Y-values appear in the attractors of the CA. Fig.3 (triangles) shows the results of such analyses. To compute this figure, we set $p = 5$, i.e. predators getting no prey loose the average of what they would gain (between 1 and 9) if they found prey in the neighbourhood; analogously: prey meeting no predators gain the average of what they would loose (between 1 and 9) if predators emerge in the neighbourhood. One clearly sees there a much more frequent appearance of prime prey cycles (dark symbols), especially with periods around 17, as compared to non-prime prey cycles (grey symbols, all being very close to the abscissa). This re-
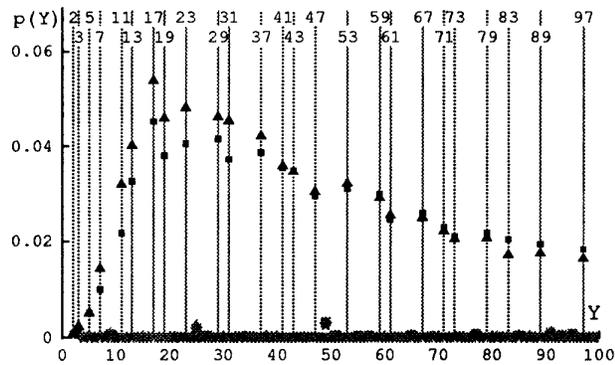
FIG. 3. Probabilities $P(Y)$ that prey periods Y are selected after the cellular automaton described in the text has converged to an attractor (triangles). At $t = 0$, cycle lengths are randomly chosen within $2 \leq X, Y \leq 100$ and randomly distributed among 10×10 cells. 10000 different initial spatial configurations are evaluated. Symbols corresponding to prime, resp. non-prime, Y-values are shown in black, resp, grey. Squares: modified model, so that $f_x(t) = +1$ if the predator finds prey, independently of the number ν of prey-populated neighbouring cells (satiation effect) and $f_y(t) = -1$ if prey meets predators, no matter how many.

sult is robust to a drastic variation of the model, this variation being displayed by squares in Fig. 3. Note that in the spatiotemporal model leading to this figure, we only restrict populations by upper bounds ($2 \leq X, Y \leq U$; U=100 in the case of Fig. 3),in contrast to the purely temporal model, in which we restricted populations to the rectangle H.

We found that changing the upper bounds U from 100 to 50 or 200 preserves the peak-like shape of $P(Y)$ vs. Y with a maximum at $Y = 17$. The peak is insensitive also to p. Furthermore, we found that the grid size does not substantially affect the results, although there is an influence on the steepness of the peaks. In fact, a 20×20 grid yields a maximum of $P(Y)$ at $Y = 13$, dropping to one fourth at $Y = 29$, as compared to the less steep peak in Fig. 3 (10×10 grid). However, results are drastically changed if the grid is too small: a 5×5 grid causes a substantial selection of non-prime Y-values.

Qualitatively, the peak-shape illustrated in Fig. 3 is explained as follows. Prey with sufficiently small Y are exposed to predators having periods $jY \leq U$, $j = 1, 2, 3, \ldots$; the number of such predators decreases on increasing Y, roughly explaining the left branch of the P(Y)-peak. In order to explain the right branch of the peak, let us compare a large prime Y (say Y=47 in Fig. 3) with a medium prime (say Y=13). As the CA proceeds in time, the prey with Y=47 will favour the survival of one predator in its neighbourhood, namely that having X=47, since only this predator can feed well on this prey; however, this predator will meet the prey in each prey generation and will thus strongly decimate the prey; more precisely: $F_y \approx -\nu$. The prey with the smaller Y=13 will favour three types of predators in its neighbourhood, namely those with X=13, 26 and 39. If we make the simplification that that these three are equally distributed, i.e. each has an average number $\nu/3$, and if we consider that X=26 only hits the prey every second prey generation, and X=39 only every third prey generation, then the fitness of the prey (which is evaluated per prey generation) is $F_y \approx -\nu/3(1 + 1/2 + 1/3) = -\nu(11/18)$. Thus the prey with Y=13 is fitter than that with Y=47, explaining the right side of the peak. A more formal explanation of the peak, including the particular value of the maximum around Y=17, is left to future research.

5   Final remark

Our results, both from the purely temporal and from the spatiotemporal model, suggest that there are generic properties of this type of dynamics that favour prime numbers. Although there are traditional methods for prime number detection [16] that are faster than the methods presented here, it is remarkable that prime-generating algorithms can be based on biological ideas.

Acknowledgements

inspiring discussions.

References

[1] May, R.M. Nature. **277**, 347-349 (1979).

[2] Murray, J.D. *Mathematical Biology.* (Springer, Berlin, 1989)

[3] Karban, R. Nature. **287**, 326-327 (1980).

[4] Kritsky, G. Nature. **341**, 288 (1989).

[5] Simon, C. Martin, A. Nature. **341**, 288-289 (1989).

[6] Hoppensteadt, F.C., Keller, J.B. Science **194**, 335-337 (1976).

[7] Bulmer, M.G. Am. Nat. **111**, 1099-1117 (1977).

[8] LLoyd,M. , Dybas, H.S. Evolution. **20**, 466-505 (1966).

[9] Berlekamp, E., Conway, J., Guy, R. *Winning Ways,* Vol.2. (Academic Press, New York, 1982) ; W. Poundstone. *The Recursive Universe.* (Oxford Univ. Press, 1985)

[10] Hassell, M.P., Comins, H.N., May, R.M. Nature **353**, 255-258 (1991).

[11] Boerlijst, M.C., Hogeweg, P. Physica **D48**, 17-28 (1991).

[12] *Nonlinear Wave Processes in Excitable Media.* Edited by Holden, A.V., Markus, M., Othmer, H.G. (Plenum, New York, 1991)

[13] Markus, M., Hess, B. Nature. **347**, 56-58 (1990).

[14] Markus, M., Nagy-Ungvarai, Zs., Hess, B. Science. **257**, 225-227 (1992).

[15] Markus, M., Kloss, G., Kusch, I. Nature. **371**, 402-404 (1994).

[16] Ribenboim, P. *The New Book of Prime Number Records.* (Springer, New York, 1991); Bressoud, D.M. *Factorization and Primality Testing.* (Springer, New York, 1989).

[17] Nowak, M.A., May, R.M. Nature. **359**, 826-829 (1992).

[18] Markus, M. Biomed. Biochim. Acta **49**, 681-696 (1990).

[19] Schepers, H., Markus, M. Physica. **A188**, 337-343 (1992).

[20] Kusch, I., Markus, M. J. theor. Biol. **178**, 333-340 (1996).

[21] Markus, M., Czajka, A., Boehm, D., Hahn, T., Schulte, T., Ribeiro, A. In: *Cellular Automata and Complex Systems.* Edited by E. Goles and S. Martinez, pp. 55-105 (1999).