Chaos in learning a simple two-person game

Yuzuru Sato[†‡], Eizo Akiyama[§], and J. Doyne Farmer[¶]

[†]Brain Science Institute, The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako, Saitama 351-0198, Japan; [§]Institute of Policy and Planning Sciences, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8573, Japan; and [¶]McKinsey Professor, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501

Communicated by Brian Skyrms, University of California, Irvine, CA, February 12, 2002 (received for review December 3, 2001)

We investigate the problem of learning to play the game of rock–paper–scissors. Each player attempts to improve her/his average score by adjusting the frequency of the three possible responses, using reinforcement learning. For the zero sum game the learning process displays Hamiltonian chaos. Thus, the learning trajectory can be simple or complex, depending on initial conditions. We also investigate the non-zero sum case and show that it can give rise to chaotic transients. This is, to our knowledge, the first demonstration of Hamiltonian chaos in learning a basic two-person game, extending earlier findings of chaotic attractors in dissipative systems. As we argue here, chaos provides an important self-consistency condition for determining when players will learn to behave as though they were fully rational. That chaos can occur in learning a simple game indicates one should use caution in assuming real people will learn to play a game according to a Nash equilibrium strategy.

Learning in Games

Most work in game theory and economics involves the assumption of perfect rationality. In this case, it is natural to characterize a game in terms of its Nash equilibria, at which neither player can achieve better performance by modifying her/his strategy. Under the more realistic assumption that the players are only boundedly rational and must learn their strategies, everything becomes more complicated. Under-learning the strategies may fail to converge to a Nash equilibrium (1) or may even be chaotic (2). Thus, understanding the learning dynamics is essential (3). Here we give an example of an elementary two-person game in which a standard learning procedure leads to Hamiltonian chaos. This example extends earlier work finding chaotic attractors in dissipative game dynamics (4). We argue that chaos is a necessary condition for intelligent adaptive players to fail to converge to a Nash equilibrium (for some related work, see ref. 5).

A good example is the game of rock–paper–scissors: rock beats scissors, paper beats rock, scissors beats paper. With possible relabelings of the three possible moves, such as "earwig–man–elephant," this ancient game is played throughout the world (6). To allow players to use their "skill," it is often played repeatedly. In contrast, two-game theorists who practice what they preach would play with the skill-free Nash equilibrium mixed strategy, which is to choose the three possible moves randomly with equal probability. (In game theory, a mixed strategy is a random combination of the pure strategies—here, rock, paper, and scissors.) On average, the Nash equilibrium mixed strategy for rock–paper–scissors has the advantage that no strategy can beat it but it also has the disadvantage that there is no strategy that it can beat. An inspection of the World Rock–Paper–Scissors Society web site (http://www.worldrps.com/gbasics.html) suggests that members of this society do not play the Nash equilibrium strategy. Instead, they use psychology to try to anticipate the moves of the other player or particular sequences of moves to try to induce responses in the other player. At least for this game, it seems that real people do not learn to act like the rational agents studied in standard game theory.

A failure to converge to a Nash equilibrium under learning can happen, for example, because the dynamics of the trajectories of the evolving strategies in the space of possibilities are chaotic. Chaos has been observed in games with spatial interactions (7) or in games based on the single-population replicator equation (4, 8). In the latter examples, players are drawn from a single population and the game is repeated only in a statistical sense, i.e., the players' identities change in repeated trials of the game.

The example we present here demonstrates Hamiltonian chaos in a two-person game, in which each player learns her/his own strategy. We observe this for a zero-sum game, i.e., one in which one player's win is always the other's loss. The observation of chaos is particularly striking because of the simplicity of the game. Because of the zero-sum condition the learning dynamics have a conserved quantity with a Hamiltonian structure (9) similar to that of physical problems, such as celestial mechanics. There are no attractors, and trajectories do not approach the Nash equilibrium. Because of the Hamiltonian structure, the chaos is particularly complex, with chaotic orbits finely interwoven between regular orbits; for an arbitrary initial condition it is impossible to say *a priori* which type of behavior will result. When the zero-sum condition is violated we observe other complicated dynamical behaviors, such as heteroclinic orbits with chaotic transients. As discussed in the conclusions, the presence of chaos is important because it implies that it is not trivial to anticipate the behavior of the other player. Thus, under chaotic learning dynamics even intelligent adaptive agents may fail to converge to a Nash equilibrium.

The Model of Learning Players

We investigate a game involving two players. At each move the first player chooses from one of m possible pure strategies (moves) with frequency $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, and similarly the second player chooses from one of n possible pure strategies with frequency $\mathbf{y} = (y_1, y_2, \ldots, y_m)$. The players update $\mathbf{x}$ and $\mathbf{y}$ based on past experience by using reinforcement learning. Behaviors that have been successful are reinforced, and those that have been unsuccessful are repressed. In the continuous time limit where the change in $\mathbf{x}$ and $\mathbf{y}$ on any given time step goes to zero under some plausible assumptions, it is possible to show (11) that reinforcement learning dynamics are described by the coupled replicator equations (see *Notes*) of the form

$$\dot{x}_i = x_i[(Ay)_i - xAy](i = 1, \ldots, n), \quad [1]$$

$$\dot{y} = y[(Bx)_j - yBx](i = 1, \ldots, m), \quad [2]$$

where A and B are the payoff matrices for the first and second players, respectively.

The relation of these equations to reinforcement learning is very intuitive. Consider the first equation, which describes the updating of the strategies of the first player: The frequency of strategy i increases proportional to [current frequency (x_i)] times [average performance relative to the mean]. $(A\mathbf{y})_i$ is the performance of strategy i (averaged over the second player's possible

[‡]To whom reprint requests should be addressed. E-mail: ysato@bdc.brain.riken.go.jp.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

www.pnas.org/cgi/doi/10.1073/pnas.032086299

moves), and $\mathbf{x}A\mathbf{y}$ is the performance averaged over all m strategies of the first player. The second equation is similar.

We investigate the dynamics of a generalized rock–paper–scissors game whose payoff matrices are

$$A = \begin{bmatrix} \varepsilon_x & -1 & 1 \\ 1 & \varepsilon_x & -1 \\ -1 & 1 & \varepsilon_x \end{bmatrix}, B = \begin{bmatrix} \varepsilon_y & -1 & 1 \\ 1 & \varepsilon_y & -1 \\ -1 & 1 & \varepsilon_y \end{bmatrix}, \quad [3]$$

where $-1 < \varepsilon_x < 1$ and $-1 < \varepsilon_y < 1$ are the payoffs when there is a tie. We place these bounds on ε because when they are violated, the behavior under ties dominates, and this more closely resembles a matching-pennies-type game with three strategies. We have placed the columns in the order "rock," "paper," and "scissors." For example, reading down the first column of **A**, in the case that the opponent plays "rock," we see that the payoff for using the pure strategy "rock" is ε_x, "paper" is 1, and "scissors" is -1.

The rock–paper–scissors game exemplifies a class of games where no strategy is dominant and no pure-strategy Nash equilibrium exists (any pure strategy is vulnerable to another). An example of a possible application is two broadcasting companies competing for the same time slot when preferences of the audience are context-dependent. Suppose, for example, that the audience prefers sports to news, news to drama, and drama to sports. If each broadcasting company must commit to their schedule without knowing that of their competitor, then the resulting game is of this type.

We consider the general case that a tie is not equivalent for both players, i.e., $\varepsilon_x \neq \varepsilon_y$. In the example above, this symmetry would be true if the audience believes that within any given category one company's programming is superior to the other. If the size of the audience is fixed, so that one company's gain is the other's loss, this is a zero-sum game corresponding to the condition $\varepsilon_x = -\varepsilon_y = \varepsilon$. In general, Eqs. **1** and **2** form a conservative system, which cannot have an attractor. If in addition $A = -B^t$, it is known that the dynamics are Hamiltonian (9). This is a stronger condition, as it implies the full dynamical structure of classical mechanics, with pairwise conjugate coordinates obeying Liouville's theorem.

Dynamical Behavior of the System

To visualize the behavior of this system, in Fig. 1 we show Poincaré sections of Eqs. **1** and **2** with the initial conditions $(x_1, x_2, x_3, y_1, y_2, y_3) = (0.5, 0.01k, 0.5 - 0.01k, 0.5, 0.25, 0.25)$ with $k = 1, 2, \ldots, 25$. This is a sample of points where the trajectories intersect the hyperplane $x_2 - x_1 + y_2 - y_1 = 0$. When $\varepsilon = 0$, our simulation indicates that the system is integrable, and trajectories are confined to quasi-periodic tori. When $\varepsilon > 0$, however, this is no longer guaranteed. As we vary ε from 0 to 0.5 without changing initial conditions, some tori collapse and become chaotic, and the trajectories cover a larger region of the strategy space. Regular and chaotic trajectories are finely interwoven; for typical behavior of this type, there is a regular orbit arbitrarily close to any chaotic orbit (12).

To demonstrate that these trajectories are indeed chaotic, we numerically compute Lyapunov exponents, which can be viewed as generalizations of eigenvalues that remain well defined for chaotic dynamics. Positive values indicate directions of average local exponential expansion, and negative values indicate local exponential contraction. Some examples are given in Table 1. The largest Lyapunov exponents are clearly positive for the first three initial conditions when $\varepsilon = 0.25$ and for the first four initial conditions when $\varepsilon = 0.5$. An indication of the accuracy of these computations can be obtained by comparing to known cases: because of the conservation condition the four exponents always sum to zero; because of the special nature of motion along trajectories plus the Hamiltonian condition the second and third

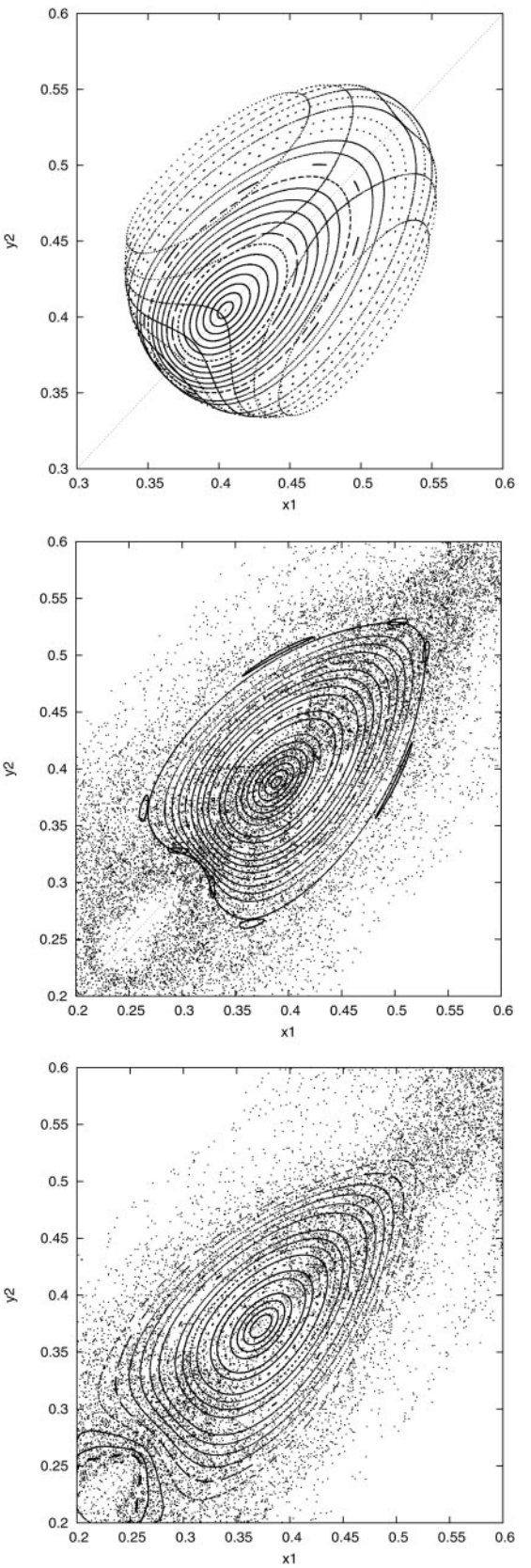

Fig. 1. Poincaré section at $x_2 - x_1 + y_2 - y_1 = 0$. Nonlinear parameters are $\varepsilon = 0$ (*Top*), $\varepsilon = 0.25$ (*Middle*), and $\varepsilon = 0.50$ (*Bottom*). The horizontal and vertical axis are x_1, y_2, respectively. Initial conditions are given as $(x_1,x_2,x_3,y_1,y_2,y_3) = (0.5,0.01k,0.5 - 0.01k,0.5,0.25,0.25)$ with $k = 1, 2, \ldots, 25$. We used a fourth-order symplectic integrator (10) for the canonical form of Hamiltonian (see *Notes*).

Table 1. Lyapunov spectra for different initial conditions (columns) and different values of the tie breaking parameter ε

ε	λ	$k = 1$	2	3	4	5
0	λ_1	+1.0	+1.4	+0.4	+0.4	+0.4
	λ_2	+0.2	+0.3	+0.3	+0.3	+0.3
	λ_3	−0.5	−0.4	−0.3	−0.3	−0.3
	λ_4	−0.7	−1.2	−0.4	−0.4	−0.4
0.25	λ_1	**+49.0**	**+35.3**	**+16.6**	+0.4	+0.4
	λ_2	+0.3	+0.3	+0.4	+0.2	+0.3
	λ_3	−0.3	−0.1	−0.4	−0.2	−0.3
	λ_4	−49.0	−35.5	−16.5	−0.4	−0.4
0.50	λ_1	**+61.6**	**+35.0**	**+28.1**	**+12.1**	+0.2
	λ_2	+0.6	+0.3	+0.1	+0.0	+0.2
	λ_3	−0.6	−0.4	−0.2	−0.1	−0.2
	λ_4	−61.5	−35.8	−28.0	−12.2	−0.3

$k = 1, 2, 3, 4, 5$ correspond to the initial conditions $(x_1, x_2, x_3, y_1, y_2, y_3) = (0.5, 0.01k, 0.5 - 0.01k, 0.5, 0.25, 0.25)$ with $k = 1, 2, \ldots, 5$. The Lyapunov exponents are multiplied by 10^3. Note that $\lambda_2 \simeq 0.0$, $\lambda_3 \simeq 0.0$, and $\lambda_4 \simeq -\lambda_1$ as expected. The Lyapunov exponents indicating chaos are shown in boldface.

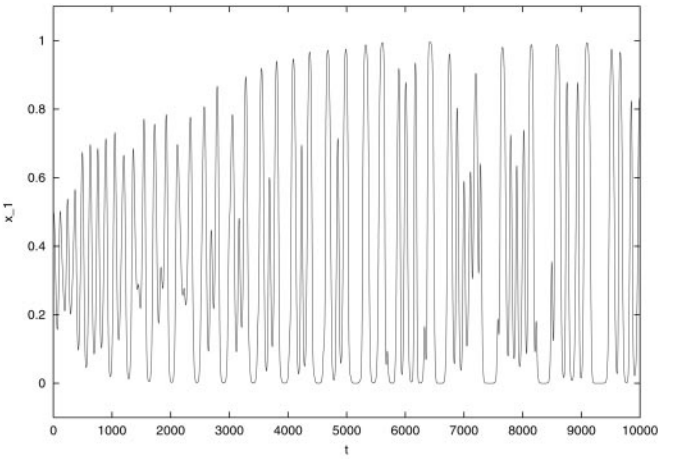

Fig. 3. The frequency of "rock" vs. time with $\varepsilon_x + \varepsilon_y > 0$ ($\varepsilon_x = 0.1$, $\varepsilon_y = -0.05$). The trajectory is a chaotic transient attracting to a heteroclinic orbit at the boundary of the simplex. The time spent near pure strategies still increases linearly on average but changes irregularly.

are always zero; and when $\varepsilon = 0$, because the motion is integrable, all Lyapunov exponents are exactly zero.

As mentioned already, this game has a unique Nash equilibrium when all responses are equally likely, i.e., $x_1^* = x_2^* = x_3^* = y_1^* = y_2^* = y_3^* = 1/3$. It is possible to show that all trajectories have the same payoff as the Nash equilibrium on average (13). However, there are significant deviations from this payoff on any given step, which are larger than those of the Nash equilibrium. Thus, a risk averse agent would prefer the Nash equilibrium to a chaotic orbit.

The behavior of the non-zero sum game is also interesting and unusual. When $\varepsilon_x + \varepsilon_y < 0$ (e.g., $\varepsilon_x = -0.1$, $\varepsilon_y = 0.05$), the motion approaches a heteroclinic cycle, as shown in Fig. 2.

Players switch between pure strategies in the order *rock* → *paper* → *scissors*. The time spent near each pure strategy increases linearly with time. This dynamics is in contrast to analogous behavior in the standard single-population replicator model, which increases exponentially with time. When $\varepsilon_x + \varepsilon_y > 0$ (e.g., $\varepsilon_x = 0.1$, $\varepsilon_y = -0.05$), as shown in Fig. 3, the behavior is similar, except that the time spent near each pure strategy varies irregularly. The orbit is an infinitely persistent chaotic transient (14).

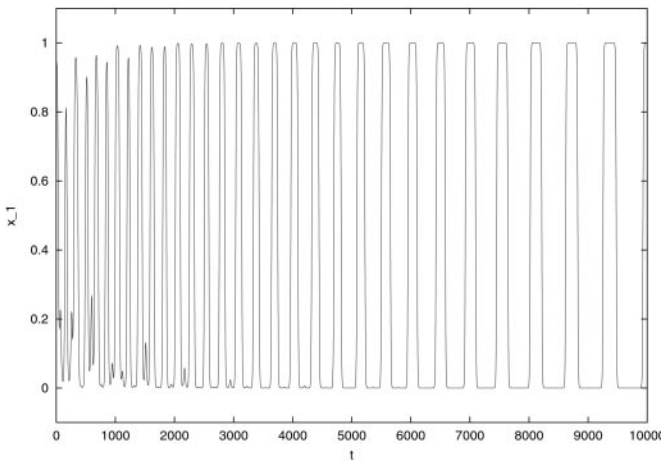

Fig. 2. The frequency of the pure strategy "rock" vs. time with $\varepsilon_x + \varepsilon_y < 0$ ($\varepsilon_x = -0.1$, $\varepsilon_y = 0.05$). The trajectory is attracted to a heteroclinic cycle at the boundary of the simplex. The duration of the intervals spent near each pure strategy increases linearly with time.

Chaos in Learning and Rationality

The emergence of chaos in learning in such a simple game illustrates that rationality may be an unrealistic approximation even in elementary settings. Chaos provides an important self-consistency condition. When the learning of her/his opponent is regular, any agent with even a crude ability to extrapolate can exploit this to improve performance. Nonchaotic learning trajectories are symptomatic that the learning algorithm is too crude to represent the behavior of a human agent. When the behavior is chaotic, however, extrapolation is difficult, even for intelligent humans. Hamiltonian chaos is particularly complex, because of the lack of attractors and the fine interweaving of regular and irregular motion. This situation is compounded for high dimensional chaotic behavior, because of the "curse of dimensionality" (15). In dimensions greater than about five, the amount of data an "econometric" agent would need to collect to build a reasonable model to extrapolate the learning behavior of her/his opponent becomes enormous. For games with more players it is possible to extend the replicator framework to systems of arbitrary dimension (Y.S. and J. P. Crutchfield, unpublished observations). It is striking that low dimensional chaos can occur even in a game as simple as the one we study here. The phase space of this game is four dimensional, which is the lowest dimension in which continuous dynamics can give rise to Hamiltonian chaos. In more complicated games in higher dimensional-state spaces we expect that chaos becomes even more common.

Many economists have noted the lack of any compelling account of how agents might learn to play a Nash equilibrium (16). Our results strongly reinforce this concern (see *Notes*), in a game simple enough for children to play. That chaos can occur in learning such a simple game indicates that one should use caution in assuming that real people will learn to play a game according to a Nash equilibrium strategy.

Notes

Coupled Replicator Equations. Eqs. **1** and **2** have the same form as the multipopulation replicator equation (17) or the asymmetric game dynamics (18), which is a model of a two-population ecology. The difference from the standard single-population replicator equation (19) is in the cross term of the averaged

performance. It has been known for some time that chaos occurs in single-population replicator equations (4, 8). This model is applicable to game theory in the specialized context where both players are forced to use the same strategy, for example, when two statistically identical players are repeatedly drawn from the same population. Here we study the context of actually playing a game, i.e., two fixed players who evolve their strategies independently.

The Hamiltonian Structure. To see the Hamiltonian structure of Eqs. **1** and **2** with **3**, it helps to transform coordinates. (**x**,**y**) exist in a six-dimensional space, constrained to a four-dimensional simplex because of the conditions that the set of probabilities **x** and **y** each sum to 1. For $\varepsilon_x = -\varepsilon_y$ we can make a transformation from $\mathbf{U} = (\mathbf{u},\mathbf{v})$ in $\mathbf{R}^4$ with $\mathbf{u} = (u_1,u_2)$ and $\mathbf{v} = (v_1,v_2)$ such as $u_i = \log \frac{x_{i+1}}{x_1}$, $v_i = \log \frac{y_{i+1}}{y_1}$ $(i = 1,2)$. The Hamiltonian is

$$H = -\tfrac{1}{3}(u_1 + u_2 + v_1 + v_2) + \log(1 + e^{u_1} + e^{u_2})(1 + e^{v_1} + e^{v_2}) \quad [4]$$

$$\dot{U} = J\nabla_U H, \quad [5]$$

(see ref. 9) where the Poisson structure J is here given as

$$J = \begin{bmatrix} 0 & 0 & 2\varepsilon & 3+\varepsilon \\ 0 & 0 & -3+\varepsilon & 2\varepsilon \\ -2\varepsilon & 3-\varepsilon & 0 & 0 \\ -3-\varepsilon & -2\varepsilon & 0 & 0 \end{bmatrix}. \quad [6]$$

We can transform to canonical coordinates

$$\dot{U}' = S\nabla_{U'} H,\ S = \begin{bmatrix} O & I \\ -I & O \end{bmatrix} \quad [7]$$

by applying the linear transformation $\mathbf{U}' = \mathbf{M}\mathbf{U}$

$$M = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\dfrac{2\varepsilon}{3(\varepsilon^2+3)} & \dfrac{\varepsilon+3}{3(\varepsilon^2+3)} & 0 & 0 \\ \dfrac{\varepsilon-3}{3(\varepsilon^2+3)} & -\dfrac{2\varepsilon}{3(\varepsilon^2+3)} & 0 & 0 \end{bmatrix} \quad [8]$$

to the Hamiltonian form (5).

Conjecture on Learning Dynamics. When regular motion occurs, if one player suddenly acquires the ability to extrapolate and the other does not, the first player's score will improve. If both players can extrapolate, it is not clear what will happen. Our conjecture is that sufficiently sophisticated learning algorithms will result either in convergence to the Nash equilibrium or in chaotic dynamics. In the case of chaotic dynamics, it is impossible for players to improve their performance because trajectories become effectively unforecastable, and in the case of Nash equilibrium, it is also impossible by definition.

We thank Sam Bowles, Jim Crutchfield, Mamoru Kaneko, Paolo Patelli, Cosma Shalizi, Spyros Skouras, Isa Spoonheim, Jun Tani, and Eduardo Boolean of the World Rock–Paper–Scissors Society for useful discussions. This work was supported by the Special Postdoctoral Researchers Program at The Institute of Physical and Chemical Research (Japan); and by grants from the McKinsey Corporation, Bob Maxfield, Credit Suisse, and Bill Miller.

1. Shapley, L. (1964) *Ann. Math. Studies* **5,** 1–28.
2. Cowen, S. (1992) Doctoral Dissertation (Univ. of California, Berkeley).
3. Akiyama, E. & Kaneko, K. (2000) *Physica* **D147,** 221–258.
4. Skyrms, B. (1996) in *The Dynamics of Norms*, eds. Bicchieri, C., Jeffrey, R. & Skyrms, B. (Cambridge Univ. Press, Cambridge, U.K.), pp. 199–222.
5. Young, P. & Foster, D. P. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 12848–12853.
6. Opie, I. & Opie, P. (1969) *Children's Games in Street and Playground* (Oxford Univ. Press, Oxford).
7. Nowak, M. A. & May, R. M. (1992) *Nature (London)* **359,** 826–829.
8. Nowak, M. A. & Sigmund, K. (1992) *Proc. Natl. Acad. Sci. USA* **90,** 5091–5094.
9. Hofbauer, J. (1996) *J. Math. Biol.* **34,** 675–688.
10. Yoshida, H. (1990) *Phys. Lett. A* **150,** 262–268.
11. Borgers, T. & Sarin, R. (1997) *J. Econ. Theory* **77,** 1–14.
12. Lichtenberg, A. J. & Lieberman, M. A. (1983) *Regular and Stochastic Motion* (Springer, New York).
13. Schuster, P., Sigmund, K., Hofbauer, J. & Wolff, R. (1981) *Biol. Cybern.* **40,** 1–8.
14. Chawanya, T. (1995) *Prog. Theor. Phys.* **94,** 163–179.
15. Farmer, J. D. & Sidorowich, J. J. (1987) *Phys. Rev. Lett.* **59,** 845–848.
16. Kreps, D. M. (1990) *Game Theory and Economic Modelling* (Oxford Univ. Press, Oxford).
17. Taylor, P. D. (1979) *J. Appl. Probability* **16,** 76–83.
18. Hofbauer, J. & Sigmund, K. (1988) *The Theory of Evolution and Dynamical Systems* (Cambridge Univ. Press, Cambridge, U.K.).
19. Taylor, P. D. & Jonker, L. B. (1978) *Math. Biosci.* **40,** 145–156.