Chaos in learning a simple two person game

Yuzuru Sato*
Brain Science Institute, RIKEN, 2-1 Hirosawa, Saitama 351, Japan

Eizo Akiyama†
Institute of Policy and Planning Sciences, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8573, Japan

J. Doyne Farmer‡
McKinsey Professor, Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501

(Dated: September 7, 2001)

We investigate the problem of learning to play a generalized rock-paper-scissors game. Each player attempts to improve her average score by adjusting the frequency of the three possible responses. For the zero-sum case the learning process displays Hamiltonian chaos. The learning trajectory can be simple or complex, depending on initial conditions. For the non-zero-sum case it shows chaotic transients. This is the first demonstration of chaotic behavior for learning in a basic two person game. As we argue here, chaos provides an important self-consistency condition for determining when adaptive players will learn to behave as though they were fully rational.

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 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], in which case understanding their dynamics is essential [2]. Here we give the first example of an elementary two person game in which a standard learning procedure leads to chaos, and argue that chaos is a necessary condition for intelligent players (with the ability to extrapolate) to fail to converge to a Nash equilibrium.

A good example is the game of rock-paper-scissors: rock beats scissors, paper beats rocks, scissors beats paper. With possible relabelings of the three possible moves, such as "earwig-man-elephant", this ancient game is played throughout the world [3]. To allow players to use their "skill", it is often played repeatedly. In contrast, two game theorists who practice what they preach will play using 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 disadvantage that there is no strategy that it can beat. An inspection of the World Rock-Paper-Scissors Society website [4] 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 appears that real people do not learn to act like rational agents in 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. This has been observed in games with spatial interactions [5], or in games based on the single population replicator equation [6, 7]. In the latter examples players are drawn from a single population and the game is repeated only in a statistical sense (i.e. the identity of the players changes in repeated trials of the game).

The example we present here is the first to demonstrate chaos in a two person game, in which each player learns her 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 this game. Because of the zero sum condition the learning dynamics have a conserved quantity, with a Hamiltonian structure 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 behav-
tors, such as heteroclinic orbits or 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 adapting agents may fail to converge to a Nash equilibrium.

We investigate a game involving two players. At each move the first player chooses from one of $m$ possible pure strategies (moves) with frequency $x = (x_1, x_2, \ldots, x_m)$, and similarly the second player chooses from one of $n$ possible pure strategies with frequency $y = (y_1, y_2, \ldots, y_n)$. The players update $x$ and $y$ based on past experience 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 each player's strategy is small, the payoff matrix $A$ and $B$ can be averaged over many repetitions of the game. The players update their strategies based on past experience. The frequency of strategy $i$ is denoted by $x_i$, and the performance of strategy $i$ is $x_i [(A y)_i - x A y]$, where $(A y)_i$ is the performance of strategy $i$ averaged over the second player's possible moves, and $x A y$ is the performance averaged over all $m$ strategies of the first player. The second equation is similar.

It has been known for some time that chaos occurs in single population replicator equations [6]. This is applicable to game theory in the specialized context where players update their strategies with frequency $x_i$ based on past experience. In the continuous time limit where the change in each player's strategy is small, the payoff matrix $A$ and $B$ can be averaged over many repetitions of the game. The players update their strategies based on past experience. The frequency of strategy $i$ is denoted by $x_i$, and the performance of strategy $i$ is $x_i [(A y)_i - x A y]$, where $(A y)_i$ is the performance of strategy $i$ averaged over the second player's possible moves, and $x A y$ 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

$$A = \begin{bmatrix}
\epsilon_x & -1 & 1 \\
1 & \epsilon_x & -1 \\
-1 & 1 & \epsilon_x
\end{bmatrix}, \quad B = \begin{bmatrix}
\epsilon_y & -1 & 1 \\
1 & \epsilon_y & -1 \\
-1 & 1 & \epsilon_y
\end{bmatrix},$$

where $-1 \leq \epsilon_x \leq 1$ and $-1 \leq \epsilon_y \leq 1$ are the payoffs when there is a tie [11]. 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 $\epsilon_x$, “paper” is 1, and “scissors” is $-1$.

The rock-paper-scissors game exemplifies a fundamental and important 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. $\epsilon_x \neq \epsilon_y$. In the example above this 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. This corresponds to the condition $\epsilon_x = -\epsilon_y = \epsilon$. In this case Eqs. (1) and (2) form a conservation system, which cannot have an attractor. Since in this case $A_{ij} = -B_{ji}$, it is known that the dynamics are Hamiltonian [12]. This is a stronger condition, as it implies the full dynamical structure of classical mechanics, with pairwise conjugate coordinates obeying Liouville’s theorem (which states that the area in phase space enclosed by each pair of conjugate coordinates is preserved).

To see the Hamiltonian structure it helps to transform coordinates. $(x, y)$ exist in a six dimensional space, constrained to a 4-dimensional simplex due to the conditions that the set of probabilities $x$ and $y$ each sum to one. For $\epsilon_x = -\epsilon_y$ we can make a transformation from $U = (u, v)$ in $R^4$ with $u = (u_1, u_2)$ and $v = (v_1, v_2)$ such as $u_i = \log \frac{x_i + 1}{x_i}$, $v_i = \log \frac{y_i + 1}{y_i}$ ($i = 1, 2$). The Hamiltonian is

$$H = -1/3(u_1 + u_2 + v_1 + v_2) + \log(1 + e^{u_1} + e^{v_2})(1 + e^{v_1} + e^{v_2})$$

$$\dot{U} = J \nabla_u H$$

where the Poisson structure $J$ is given as

$$J = \begin{bmatrix}
0 & 0 & 2\epsilon & 3 + \epsilon & 0 & 0 \\
0 & 0 & -3 - \epsilon & 2\epsilon & 0 & 0 \\
-2\epsilon & 3 - \epsilon & 0 & 0 & 0 & 0 \\
-3 - \epsilon & -2\epsilon & 0 & 0 & 0 & 0
\end{bmatrix}.$$
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 I. The largest Lyapunov exponents are clearly positive for the first three initial conditions when \( \epsilon = 0.25 \) and for the first four initial conditions when \( \epsilon = 0.5 \). 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 are always zero, and when \( \epsilon = 0 \), because the motion is integrable, all Lyapunov exponents are exactly zero.

<table>
<thead>
<tr>
<th>( \epsilon )</th>
<th>( \lambda )</th>
<th>( k=1 )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( \lambda_1 )</td>
<td>+1.0</td>
<td>+1.4</td>
<td>+0.4</td>
<td>+0.4</td>
<td>+0.4</td>
</tr>
<tr>
<td></td>
<td>( \lambda_2 )</td>
<td>+0.2</td>
<td>+0.3</td>
<td>+0.3</td>
<td>+0.3</td>
<td>+0.3</td>
</tr>
<tr>
<td></td>
<td>( \lambda_3 )</td>
<td>−0.5</td>
<td>−0.4</td>
<td>−0.3</td>
<td>−0.3</td>
<td>−0.3</td>
</tr>
<tr>
<td></td>
<td>( \lambda_4 )</td>
<td>−0.7</td>
<td>−1.2</td>
<td>−0.4</td>
<td>−0.4</td>
<td>−0.4</td>
</tr>
<tr>
<td>0.25</td>
<td>( \lambda_1 )</td>
<td>+49.0</td>
<td>+35.3</td>
<td>+16.6</td>
<td>+0.4</td>
<td>+0.4</td>
</tr>
<tr>
<td></td>
<td>( \lambda_2 )</td>
<td>+0.3</td>
<td>+0.3</td>
<td>+0.4</td>
<td>+0.2</td>
<td>+0.3</td>
</tr>
<tr>
<td></td>
<td>( \lambda_3 )</td>
<td>−0.3</td>
<td>−0.1</td>
<td>−0.4</td>
<td>−0.2</td>
<td>−0.3</td>
</tr>
<tr>
<td></td>
<td>( \lambda_4 )</td>
<td>49.0</td>
<td>−35.5</td>
<td>−16.5</td>
<td>−0.4</td>
<td>−0.4</td>
</tr>
<tr>
<td>0.50</td>
<td>( \lambda_1 )</td>
<td>+61.6</td>
<td>+35.0</td>
<td>+28.1</td>
<td>+12.1</td>
<td>+0.2</td>
</tr>
<tr>
<td></td>
<td>( \lambda_2 )</td>
<td>+0.6</td>
<td>+0.3</td>
<td>+0.1</td>
<td>+0.0</td>
<td>+0.2</td>
</tr>
<tr>
<td></td>
<td>( \lambda_3 )</td>
<td>−0.6</td>
<td>−0.4</td>
<td>−0.2</td>
<td>−0.1</td>
<td>−0.2</td>
</tr>
<tr>
<td></td>
<td>( \lambda_4 )</td>
<td>−61.5</td>
<td>−35.8</td>
<td>−28.0</td>
<td>−12.2</td>
<td>−0.3</td>
</tr>
</tbody>
</table>

TABLE I: Lyapunov spectra for different initial conditions (columns) and different values of the tie breaking parameter \( \epsilon \). \( k = 1, 2, 3, 4, 5 \) correspond to the initial conditions \( (x_1, x_2, y_1, y_2, y_3) = (0.5, 0.01k, 0.5 - 0.01k, 0.5, 0.25, 0.25) \) 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.

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 [15]. 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 \( \epsilon_x + \epsilon_y < 0 \) (e.g. \( \epsilon_x = -0.1, \epsilon_y = 0.05 \)), the motion approaches a heteroclinic cycle, as shown in Fig.2. Players switch between pure strategies in the order \( \text{rock} \rightarrow \text{paper} \rightarrow \text{scissors} \). The time spent near each pure strategy increases linearly with time. (This is in contrast to analogous behavior in the standard single population replicator model, which increases exponentially with time.) When \( \epsilon_x + \epsilon_y > 0 \) (e.g. \( \epsilon_x = 0.05, \epsilon_y = 0.1 \)), the motion is integrable, all Lyapunov exponents are exactly zero.

FIG. 1: Poincaré section at \( x_2 - x_1 + y_2 - y_1 = 0 \): Nonlinear parameters are \( \epsilon = 0 \) (top), \( \epsilon = 0.25 \) (middle) and \( \epsilon = 0.50 \) (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 an 8th-order symplectic integrator [14].
FIG. 2: The frequency of the pure strategy “rock” vs. time with $\epsilon_x + \epsilon_y < 0$ ($\epsilon_x = -0.1$, $\epsilon_y = 0.05$). 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.

FIG. 3: The frequency of “rock” v.s. time with $\epsilon_x + \epsilon_y > 0$ ($\epsilon_x = 0.1$, $\epsilon_y = -0.05$). 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.

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 opponent is regular, any agent with even a crude ability to extrapolate can exploit this to improve performance. Non-chaotic 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 [17]. Hamiltonian systems are particularly complex, due to the lack of attractors and the fine interweaving of regular and irregular motion. This situation is compounded for high dimensional chaotic behavior, due to the “curse of dimensionality” [18]. 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 opponent becomes enormous. For games with more players it is possible to extend the replicator framework to systems of arbitrary dimension [9], which we intend to investigate in the future. It is striking that low dimensional chaos can occur even in a game as simple as the one we study here [19]. In more complicated games with higher dimensional chaos we expect that it becomes more common.

Many economists have noted the lack of any compelling account of how agents might learn to play a Nash equilibrium [20]. Our results strongly reinforce this concern, in a game simple enough for children to play. The fact 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.

REFERENCES

[11] We place these bounds on $\epsilon$ because when they are violated, the behavior under ties dominates, and this more closely resembles a matching-pennies-type game with three strategies.
We can transform to canonical coordinates
\[ \dot{\mathbf{U}} = S \nabla \mathbf{U}^t \mathbf{H}, \quad S = \begin{bmatrix} O & I \\ -I & O \end{bmatrix}. \]

by applying the linear transformation \( \mathbf{U} = \mathbf{M} \mathbf{U} \)

\[ M = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & -\frac{2c}{3(c^2 + 3)} & 0 & 1 \\ 0 & \frac{c + 3}{3(c^2 + 3)} & 0 & 0 \\ -\frac{2c}{3(c^2 + 3)} & 0 & 0 & 0 \end{bmatrix}. \]

to the Hamiltonian form (5).

When regular motion occurs, if one player suddenly acquires the ability to extrapolate and the other doesn’t, 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.

The authors would like to thank Sam Bowles, Jim Crutchfield, Mamoru Kaneko, Paolo Patelli, Cosma Shalizi, Spyros Skouras, Isa Spoonheim, Jun Tani, and Eduardo Boolean of the World RPS Society for useful discussions. This work was supported by The Special Postdoctoral Researchers Program at RIKEN, and by grants from the McKinsey Corporation, Bob Maxfield, Credit Suisse, and Bill Miller.