Learning the Optimal Strategy in a Zero-Sum Game

Vincent P. Crawford

Econometrica, Volume 42, Issue 5 (Sep., 1974), 885-891.

Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28197409%2942%3A5%3C885%3ALTOSIA%3E2.0.CO%3B2-D

Econometrica, Vol. 42, No. 5 (September, 1974)

LEARNING THE OPTIMAL STRATEGY IN A ZERO-SUM GAME

BY VINCENT P. CRAWFORD[1]

This paper investigates the possibility of arriving at the mixed-strategy solution of a zero-sum two-person game through an iterative learning process. Learning takes place during repeated play of the game, in which the players have no direct knowledge of the payoff matrix but are allowed to record what happens during play. In this context, all members of a wide class of behaviorally plausible learning mechanisms are shown to be locally unstable for "almost all" zero-sum two-person games with mixed-strategy solutions.

1. INTRODUCTION

ECONOMISTS HAVE KNOWN for a long time that finding the optimal strategy in a zero-sum two-person game is equivalent to finding the solution of an appropriately defined linear program.[2] It is less well known that there are other ways to solve zero-sum games. Von Neumann [**10**] himself and G. W. Brown [**2**] have provided analog methods that converge to the correct solution and are conceptually simpler than the method based on the Dantzig simplex algorithm, which rapidly becomes impractical for unaided humans to apply as the size of the payoff matrix increases.

Because economists occasionally apply zero-sum game theory as a descriptive model,[3] it would be of interest to find a simple, behaviorally plausible alternative to the simplex method in order to provide a stronger justification for descriptive applications. Unfortunately the papers of von Neumann and Brown, while highly successful in terms of their goals of improving the computational methods, do not provide such an alternative. Brown's algorithm, styled "solution by fictitious play," comes closer than that of von Neumann. He suggests that if two "statisticians" played the same game many times, at each play selecting the pure strategy that would have been optimal against the mixture of all his opponent's past pure strategy choices, the mixed strategies generated by this process would average in the long run to the solution of the game. This conjecture was later proved by Julia Robinson [**7**]. Brown's "statisticians," however, arrive at the solution through myopic attempts to outguess their opponents, much like the duopolists of Cournot, and thus fail to accept the rationale of the mixed-strategy solution.[4] They also give fully as much weight to the distant past as to the immediate past in estimating their opponents' current strategy mixtures.

[1] I wish to thank R. L. Bishop, F. M. Fisher, and M. Lonoff for helpful comments and suggestions, but I am responsible for any errors that remain. Financial support from the National Science Foundation is gratefully acknowledged.

[2] See, for example, Gale, Kuhn, and Tucker [**3**].

[3] Perhaps the best examples of this are in Shubik [**9**].

[4] For a discussion of this rationale, see, for example, von Neumann and Morgenstern [**11**, pp. 143–145].

In 1966, Rapoport [6, pp. 145–157] took a first step in the behavioral direction, analyzing what he called "an inductive theory of games." He considers two players repeatedly playing a zero-sum game without direct knowledge of the payoff matrix. They begin with an arbitrary mixed strategy and change it experimentally, responding in proportion to improved payoffs by continuing the change that resulted in the improvement. The system of linear differential equations that reflects this process generates sinusoidal oscillations about the solution of the game, with constant amplitude determined by initial conditions. While the solution of the game is an equilibrium in this context, it is not a stable one. Rapoport nevertheless concludes that players can arrive at the solution through empirically based revisions of their strategies, since the mixed strategies generated by his process will average over time to the solution.

This study, while of interest, has several important shortcomings. By substituting expectations for the realizations of the random payoffs, Rapoport is using a linear deterministic model to describe a stochastic process. While this simplification is often justified, it is a particularly misleading one when the deterministic model falls on the "knife-edge" between stability and instability. In this case, the deterministic model will generate oscillations with constant amplitude, while the "true" stochastic model generates oscillations whose expected amplitude increases over time.[5] Therefore, in a stochastic version of Rapoport's model the natural zero/one probability boundaries will soon come into play, and the linear model cannot continue to hold. Additional difficulties are that Rapoport's formulation is limited to the case of a 2×2 payoff matrix, and has no natural generalization to the case of larger payoff matrices, and that his players fail to use all available information in adjusting their strategies. Finally, the differential equation formulation seems inappropriate for a process that is essentially discrete.

In this light, I set out to find a simple, behaviorally plausible model of the learning process that takes place when two players repeatedly play a zero-sum game without direct knowledge of the payoff matrix. I hoped that such a process would converge in time to the solution of the game. The outcome of this research is, unfortunately, the negative result proved in this paper, that all members of a wide class of such models are locally unstable for "almost all" games without solutions in pure strategies. In Section 2 of this paper I will present the model and prove the basic theorem, which applies directly only to games whose mixed-strategy solutions have all positive probabilities. In Section 3 I will argue that the results of Section 2 also apply to "almost all" games whose mixed-strategy solutions include some zero probabilities as well. Finally, in Section 4 I will discuss some simulation experiments with the model and the implications of the result.

2. A SIMPLE MODEL OF LEARNING BEHAVIOR

Consider two players repeatedly playing a zero-sum $(N + 1) \times (M + 1)$ matrix game whose solution is in mixed strategies with all positive probabilities.[6] They have

[5] See Samuelson [8, p. 268 and 336 ff].

[6] This assumption is even more restrictive than it seems. For example, if the solution of the game is unique, the assumption that it has all positive probabilities implies that $M = N$, that is, that the payoff matrix is square.

no direct knowledge of the payoff matrix, but are allowed to record their payoff together with the pure strategy actually played after each play of the game. They pause after each group of P plays to evaluate their returns and adjust their strategy mixtures in an attempt to improve their returns if possible. P is taken to be sufficiently large for the law of large numbers to make the difference between average returns and their expectations negligible; this implies that information from still earlier periods is obsolete. It is highly unlikely that this simplification will affect the qualitative conclusions reached in this case.

Under these conditions, it is reasonable to assume that each player adjusts the probability with which he plays each pure strategy in response to the difference between the average return from that strategy and the average return from the entire game over the period just completed. To remedy the difficulty that the "probabilities" adjusted in this way need not sum to unity, while retaining symmetry in the treatment of the variables, the "probabilities" are then deflated by their sums.[7] Since we will be dealing with local variations around a point in the interior of the feasible region, the natural boundaries will not be immediately effective. Ignoring these boundaries for the present, the system of nonlinear difference equations implied by the above assumptions can be written as follows:

$$x_{it} = \frac{x_{it-1} + F[\psi^i(x_{t-1}, y_{t-1})]}{1 + \sum_{k=1}^{N+1} F[\psi^k(x_{t-1}, y_{t-1})]} \qquad (i = 1, \ldots, N+1), \tag{2.1}$$

$$y_{jt} = \frac{y_{jt-1} + G[\phi^j(x_{t-1}, y_{t-1})]}{1 + \sum_{h=1}^{M+1} G[\phi^h(x_{t-1}, y_{t-1})]} \qquad (j = 1, \ldots, M+1), \tag{2.2}$$

where $x_t = (x_{1t}, \ldots, x_{Nt})$ and $y_t = (y_{1t}, \ldots, y_{Mt})$ are the vectors of independent[8] mixed strategy probabilities chosen by players X and Y at time t; F and G are any continuous and differentiable functions[9] such that $F(0) = G(0) = 0$ and $F'(\ ) > 0$, $G'(\ ) > 0$; and

$$\psi^i(x_t, y_t) \equiv f^i(y_t) - H(x_t, y_t) \qquad (i = 1, \ldots, N+1), \tag{2.3}$$

$$\phi^j(x_t, y_t) \equiv g^j(x_t) + H(x_t, y_t) \qquad (j = 1, \ldots, M+1), \tag{2.4}$$

are functions representing the difference for each player between his expected gain from playing pure strategy i or j and his expected gain from a play of the entire game, recalling that the pure-strategy expectations, f^i and g^j, are independent of

[7] Allowing one of each player's probabilities to be determined as a residual is simpler but asymmetrical, and leads to the same ultimate conclusions.

[8] Summing (2.1) over i and (2.2) over j yields $\Sigma_{i=1}^{N+1} x_{it} = 1 = \Sigma_{j=1}^{M+1} y_{jt}$, so if all but one of the x_{it} or y_{jt} are given, the remaining one is determined.

[9] Nothing in the analysis below requires that the F and G functions be the same for each pure strategy i or j, but it seems unlikely that a reasonable player would use a different adjustment rule for each pure strategy.

how often those strategies are played, and that player X's expected gain for an entire play, $H(x_t, y_t)$, is also Y's expected loss.

Equilibrium of this system occurs when

$$\psi^i(x^*, y^*) = 0 = \phi^j(x^*, y^*) \qquad (i = 1, \ldots, N+1; j = 1, \ldots, M+1), \tag{2.5}$$

since $F(0) = G(0) = 0$; in general, it is unique. Expression (2.5) holds for all i and j when the solution of the game has all positive probabilities; otherwise, some of the boundaries are effective and some of the equations are replaced by inequalities. For the analysis of the stability of (2.1)–(2.2), we need one additional well-known property of the all-positive-probability solution:

$$\frac{\partial \psi^h(x, y^*)}{\partial x_i} = 0 = \frac{\partial \phi^k(x^*, y)}{\partial y_j}, \qquad \begin{matrix}(h = 1, \ldots, N; i = 1, \ldots, N;\\ k = 1, \ldots, M; j = 1, \ldots, M),\end{matrix} \tag{2.6}$$

where the time subscripts have been suppressed for clarity and (x^*, y^*) is the solution of the game and the equilibrium of (2.5). Expression (2.6) follows directly from partial differentiation of (2.3) and (2.4) and from the fact that, if a player's opponent is playing his optimal mixed strategy in this case, adjusting his own strategy has no effect on his expected return for the entire game since he is only changing the weights in a weighted average of a constant. Partial differentiation of (2.1) and (2.2) and substitution from (2.5) and (2.6) yields

$$\frac{\partial x_{it}}{\partial x_{it-1}} = 1 = \frac{\partial y_{jt}}{\partial y_{jt-1}}, \qquad (i = 1, \ldots, N; j = 1, \ldots, M) \tag{2.7}$$

and

$$\frac{\partial x_{it}}{\partial x_{ht-1}} = 0 = \frac{\partial y_{jt}}{\partial y_{kt-1}}, \qquad \begin{matrix}(i = 1, \ldots, N; h = 1, \ldots, N; i \neq h;\\ j = 1, \ldots, M; k = 1, \ldots, M; j \neq k),\end{matrix} \tag{2.8}$$

where the partial derivatives are evaluated at (x^*, y^*).

We can investigate the local stability of the system (2.1)–(2.2) in the positive mixed-strategy case by expanding it in a Taylor's series about its equilibrium, neglecting nonlinear terms, and working in deviations from equilibrium for ease of notation. Define $\bar{x}_t \equiv x_t - x^*$ and $\bar{y}_t \equiv y_t - y^*$, and let z' stand for the transpose of z. Then the linearized version of (2.1)–(2.2) is given by the following:

$$\begin{bmatrix} \bar{x}_t' \\ \hline \bar{y}_t' \end{bmatrix} = \left[\begin{array}{c|c} I & B \\ (N \times N) & \\ \hline C & \begin{matrix} \\ I \\ (M \times M) \end{matrix} \end{array}\right] \begin{bmatrix} \bar{x}_{t-1}' \\ \hline \bar{y}_{t-1}' \end{bmatrix}. \tag{2.9}$$

Clearly, if A is the matrix in (2.9), $\operatorname{tr} A = M + N = \sum_{i=1}^{M+N} \lambda_i$, where the λ_i are the characteristic roots of A. It is easy to verify that this implies that either all the characteristic roots of A are equal to unity or that at least one of them has a modulus greater than unity. In the latter case the system (2.9) is unstable, and thus the system (2.1)–(2.2) is locally unstable. In the former case, if the characteristic

vectors of A are not independent, the solution to (2.9) has a polynomial time trend and is again unstable. If the characteristic vectors are independent and all the characteristic roots equal unity, then $A = I$.[10] Even when all the characteristic roots equal unity, (2.1)–(2.2) is locally unstable except in the trivial case where B and C are zero matrices, which happens only on a set of measure zero. An example where this happens is the game where all the entries in the payoff matrix are the same; I do not know if it can happen in more interesting games.

We have now established the following theorem:

THEOREM 2.1: *For all but a set of measure zero discrete zero-sum two-person games with solutions in positive mixed strategies, the learning process described by equations* (2.1) *and* (2.2) *is locally unstable.*

It may be briefly noted that even if we had followed Rapoport in modeling the learning process as a system of differential equations instead of difference equations, the conclusion of Theorem 2.1 would have been the same.[11]

3. EXTENSION TO THE GENERAL CASE

Theorem 2.1 would be of some interest even if nothing could be said about the more general case where some of the components of the optimal mixed strategies are zero. It is possible to argue, however, that the results of the previous section extend to all but a set of measure zero of the games in this class as well. In the discussion that follows, degenerate games having pure strategies that are assigned zero probabilities in the solution but have expected returns at the solution equal to the value of the game are excluded from consideration; this occurs only on a set of measure zero. In the games that remain, if a pure strategy is assigned a zero probability in the solution, that pure strategy must have below-average returns throughout a neighborhood of the solution. To prove this, assume the contrary. We have already ruled out the possibility of average returns. If such a pure strategy had above-average returns at the solution, then raising its probability above zero at the expense of some pure strategy with below-average or average returns would improve the player's expected return from a play of the entire game, which contradicts a well-known property of the mixed-strategy solution. Since the functions f^i and g^j are continuous, pure strategies assigned zero probabilities in the solution must have below-average returns throughout a neighborhood of the solution as well.

It follows that players whose behavior can be described by (2.1)–(2.2) will never raise these probabilities above zero in the neighborhood of the solution, except possibly in the degenerate case excluded from consideration above. The behavior of the system (2.1)–(2.2) is independent of the entries in the rows (or columns) of

[10] See McManus [5]. This fact was pointed out to me by F. M. Fisher.

[11] In the differential equation model, there does not seem to be any simple way to deflate "probabilities" and preserve symmetry, so it may be necessary to use the residual approach mentioned in footnote 7 above.

zero-probability pure strategies in the payoff matrix, since these entries enter ψ^i or ϕ^j only with zero weights. Therefore, for our purposes, we can ignore zero-probability pure strategies, striking them from the payoff matrix without loss of generality. Theorem 2.1 then applies to the general mixed-strategy case as well, and we have established the following theorem:

THEOREM 3.1: *For all but a set of measure zero discrete zero-sum two-person games with mixed-strategy solutions, the learning process described by equations* (2.1) *and* (2.2) *is locally unstable.*

4. CONCLUSIONS

This paper has proved that a wide class of reasonable learning models is locally unstable for "almost all" zero-sum games with mixed-strategy solutions. These learning models could be used to describe behavior in zero-sum games with pure-strategy saddle points, and perhaps in non-zero-sum, n-person games as well. In both cases, the simplifying properties of mixed-strategy zero-sum games used here would be absent, so it is unlikely that any general conclusions would emerge.

The model has been simulated in the 2×2 case to determine its global properties. The results of a few simple experiments can be summarized here:

(i) The model locates pure-strategy saddle points globally, at least in these simple games.[12]

(ii) After the initial departure from mixed-strategy equilibrium (induced by the failure of random payoffs to realize their expectations exactly), the model oscillates generally "around" the equilibrium, with amplitude increasing over time. The violence of these oscillations increases with $F'(\ )$ and $G'(\ )$ as might be expected (these are constant since F and G were linear in the experiments), and for moderate values of $F'(\ )$ and $G'(\ )$ the zero/one probability boundaries become effective quite often.

(iii) In these cases, the boundaries cause a consistent distortion of the average observed mixed strategies toward the center, $x_t = [1/(N+1), \ldots, 1/(N+1)]$ and $y_t = [1/(M+1), \ldots, 1/(M+1)]$, and in general the property of Rapoport's model that the observed probabilities average over time to the optimal strategies is lost.

The hypothesis of optimizing behavior is made almost universally in microeconomic theory, both for its mathematical convenience and because it so often yields unambiguous results in comparative statics. A traditional justification for this behavioral hypothesis has been that, say, businessmen, while possibly ignorant of mathematical programming and other optimization techniques, have their own common-sense ways of finding the optimal policy. I believe that the result proved in this paper has the implication that, if zero-sum game theory is to be applied as a descriptive model of the behavior of economic agents, economists must be prepared to postulate somewhat greater sophistication on the part of these agents than has been customary.

Massachusetts Institute of Technology

Manuscript received April, 1973; last revision received March, 1974.

[12] In fact, it is possible to prove this in the 2×2 case.

REFERENCES

[1] BAUMOL, WILLIAM J.: *Economic Dynamics: an Introduction*. 3rd ed. London: The Macmillan Company, 1970.

[2] BROWN, G. W.: "Iterative Solutions of Games by Fictitious Play," in *Activity Analysis of Production and Allocation*, Cowles Commission Monograph 13, ed. T. C. Koopmans. New York: John Wiley and Sons, Inc., 1951.

[3] GALE, D., H. KUHN, AND A. TUCKER: "Linear Programming and the Theory of Games," in *Activity Analysis of Production and Allocation*, Cowles Commission Monograph 13, ed. T. C. Koopmans. New York: John Wiley and Sons, Inc., 1951.

[4] LUCE, R. DUNCAN, AND HOWARD RAIFFA: *Games and Decisions: Introduction and Critical Survey*. New York: John Wiley and Sons, Inc., 1944.

[5] MCMANUS, M.: "Dynamic Cournot-type Oligopoly Models—a Correction," *Review of Economic Studies*, 29 (1962), 337–339.

[6] RAPOPORT, ANATOL: *Two Person Game Theory: the Essential Ideas*. Ann Arbor, Michigan: The Univerity of Michigan Press, 1966.

[7] ROBINSON, JULIA: "An Iterative Method of Solving a Game," *Annals of Mathematics*, 54 (1951), 296–301.

[8] SAMUELSON, PAUL A.: *Foundations of Economic Analysis*. New York: Atheneum, 1970.

[9] SHUBIK, MARTIN: *Strategy and Market Structure*. New York: John Wiley and Sons, Inc., 1959.

[10] VON NEUMANN, JOHN: "A Numerical Method to Determine Optimum Strategy," *Naval Research Logistics Quarterly*, I (1954), 109–115.

[11] VON NEUMANN, JOHN, AND OSKAR MORGENSTERN: *Theory of Games and Economic Behavior*. New York: John Wiley and Sons, Inc., 1947.