THE LIMITS OF RECIPROCITY: SOLUTION CONCEPTS AND REACTIVE STRATEGIES IN EVOLUTIONARY EQUILIBRIUM MODELS

by

Juan Carlos Martinez Coll
Universidad de Malaga

and

Jack Hirshleifer
University of California, Los Angeles

UCLA Dept. of Economics
Working Paper #577
December, 1989

Abstract

The Nash condition (misleadingly termed "collective stability" by Axelrod) is not sufficiently strong for evolutionary equilibrium. An evolutionary equilibrium must be an attractor, that is, must have a convergency zone (under the dynamic process postulated) that consists of all the points in its neighborhood. Such an attractor may be an Evolutionary Equilibrium Point (EEP) or an Evolutionary Equilibrium Region (EER); no single point in an EER has a full convergency zone, but the region as a whole attracts all the trajectories in its neighborhood.

Equilibrium states were evaluated for round-robin tournaments in Prisoners' Dilemma and Chicken environments. The archetype strategies (COOPERATE and DEFECT in Prisoners' Dilemma, COWARD and DAREDEVIL in Chicken) were augmented by the reactive strategies TIT FOR TAT and BULLY -- entering into the competitions separately in 3x3 interactions, and jointly in 4x4 interactions. For simplicity, we employed an "instant response" assumption that stacked the deck somewhat in favor of the reactive strategies, and especially in favor of TIT FOR TAT.

A variety of different evolutionary outcomes emerged, but in no case was all-TIT FOR TAT ever an EEP. Some conditions generated an equilibrium with all-nice behaviors, TIT FOR TAT being part of a strategy mixture, but other assumptions led to all-mean equilibria or to multiple possibilities. For consistency with what is observed in the real world, a credible model should imply an equilibrium in which both mean and nice behaviors are represented. Remarkably, <u>none</u> of the conditions postulated here led to such an outcome. This unsatisfying result suggests the importance of introducing a PUNISHER strategy as described elsewhere by the authors.

LIST OF FIGURES

LIST OF TABLES

THE LIMITS OF RECIPROCITY: SOLUTION CONCEPTS AND REACTIVE STRATEGIES IN EVOLUTIONARY EQUILIBRIUM MODELS[1]

Owing mainly to Axelrod's strudies of evolutionary competition in an iterated Prisoners' Dilemma context,[2] a rather astonishing claim has come to be widely accepted: that the simple reciprocity behavior known as TIT FOR TAT is a best strategy not only in the particular environment modelled by Axelrod's simulations, but quite generally. Or more sweepingly, that natural selection of TIT FOR TAT explains the evolution of social cooperation over the whole range of life and can provide the basis for cooperation even in complex social interactions among humans.[3]

Axelrod was by no means the first to explore the evolutionary implications of reciprocity strategies.[4] But the very favorable reception of his book was well warranted in the light of the dramatic interest of the tournament competitions, the charm of the author's style, and the wisdom and breadth of knowledge displayed throughout. Nevertheless, the actual <u>analysis</u>[5] had a number of serious flaws, as pointed in an earlier paper by the present authors.[6] These flaws were of two main types:

<u>Failure to appreciate the severity of limiting assumptions:</u>

(1) Prisoners' Dilemma, while important, is but one instance within the broad range of mixed-incentive payoff environments.[7] (2) Contending agents, even if they can be regarded as playing Prisoners' Dilemma, are not ordinarily in <u>evolutionary</u> competition -- the latter being a very special form of interaction involving a multi-generation sequence of random encounters among members of a large population within a constant payoff environment.[8] (3) There are many different kinds of evolutionary competitions. One type is the round-robin tournament postulated by Axelrod; another type is the elimination

tournament. In elimination tournaments, our analysis showed, TIT FOR TAT has no survival value whatsoever.[9] (4) TIT FOR TAT requires more complex capabilities than the simpler archetype strategies that define the Prisoners' Dilemma game. A TIT FOR TAT player must be able to recognize different types of opponents, and must have correspondingly a more extended behavioral repertory. When any cost of complexity is assessed on this score, we showed that TIT FOR TAT cannot survive at all in evolutionary competition (owing to free-riding on the part of COOPERATE players).[10] Another approach is to admit some chance of error when TIT FOR TAT players attempt to identify the opponent's type. Curiously, TIT FOR TAT does somewhat better here, because mistaken identification provides a way of punishing the free-riding COOPERATE players.[11]

Incorrect equilibrium concept:

As will be shown in more detail below, Axelrod's so-called "collective stability" criterion does not, in general, have the evolutionary stability property he claimed for it.[12] In consequence he was led to make a number of technically erroneous statements. Crucial among these was the assertion that an all-TIT FOR TAT population is an evolutionarily stable outcome of the iterated Prisoners' Dilemma. Our analysis showed that, even accepting all his other assumptions, an all-TIT FOR TAT population has only vanishingly small probability of being the stable evolutionary outcome. Instead, the typical outcome is a mixture of TIT FOR TAT and COOPERATE.[13]

There is one significant difference between our model and Axelrod's, however. Our analyses, in line with the seminal work of Maynard Smith [1976, 1982], permit the TIT FOR TAT player an "instant response" capability[14] -- whereas, for Axelrod, only in the next round can a TIT FOR TAT player respond to his opponent's prior move. Since the deferred-response assumption allows a DEFECT player to "free ride" for at least one round, using our assumption instead stacks the deck somewhat in favor of TIT FOR TAT.[15]

Turning to more general issues, the search for an evolutionary corner solution in which only a single strategy (whether TIT FOR TAT or whatever) survives in the population is fundamentally misguided. In the real world a coexisting mixture of strategies, more cooperative ones together with less cooperative ones, manage simultaneously to maintain themselves in the competitive struggle for evolutionary survival. Theoretical models should be designed to generate the interior solutions consistent with this observation.

In what follows, for both Prisoners' Dilemma and Chicken payoff environments we will be allowing the archetype strategies that define these games to interact with each of the reactive strategies -- TIT FOR TAT as a "nice" reactive strategy and its "mean" counterpart BULLY. Chicken is particularly significant since it has been extensively studied by evolutionary biologists (under the alternative name Hawk-Dove),[16] and has therefore been historically associated with the development of the concept of evolutionary stability stemming from the seminal work of John Maynard Smith. In addition, we will also be presenting some results for 4x4 interactions, that is, when TIT FOR TAT and BULLY are both in play.

Apart from these substantive aims, the present paper has two main methodological objectives: (i) as already indicated, to demonstrate the correct equilibrium concept for evolutionary competitions, and in addition (ii) to explore the distribution patterns of strategies away from equilibrium. Some equilibria are approached rapidly, others only slowly or not at all. A set of populations subject to a given evolutionary process will generate darker or lighter "probability clouds" over equilibrium and non-equilibrium states.[17]

1. GAMES AND STRATEGIES

In Table I, Matrix 1 defines Prisoners' Dilemma -- the payoffs being ranked from 4 as best to 1 as worst. Matrix 2 similarly defines Chicken. In Prisoners' Dilemma the archetype strategies are the familiar COOPERATE and DEFECT options; the corresponding choices in Chicken will be termed COWARD and DAREDEVIL.[18]

In both environments each player achieves a best outcome (payoff 4) playing D against C (DEFECT against COOPERATE in Prisoners' Dilemma and DAREDEVIL against COWARD in Chicken) -- that is, if he plays "mean" against an opponent choosing "nice." And in either game each player obtains his second-best outcome (payoff 3) when both play "nice." Where Chicken and Prisoners' Dilemma differ is that in Chicken the worst outcome (payoff 1) for each player occurs when both choose the "mean" option, whereas in Prisoners' Dilemma the lowest payoff goes to the exploited member of a DEFECT-COOPERATE pair. So in Chicken the opposition of interests between the players is less total: the parties have more to gain from avoiding the mean-mean outcome.

The 3x3 interactions to be taken up first always incorporate the two archetype strategies (COOPERATE and DEFECT in Prisoners' Dilemma, COWARD and DAREDEVIL in Chicken). To these we adjoin a third "reactive" strategy -- either TIT FOR TAT or BULLY. A later section will consider the 4x4 interactions where all of these strategy options are simultaneously in competition.

A TIT FOR TAT (TFT) player, in either Prisoners' Dilemma or in Chicken, opens with a nice move and subsequently echoes the opponent's previous choice.[19] BULLY does the opposite. Consistent with the idea that real-life bullies exploit weak opponents but back away from aggressive ones, in our analysis a BULLY player opens with a mean move and subsequently does the opposite of the opponent's previous choice.[20] We will be attributing an instant-response capability to both TIT FOR TAT and BULLY: a TFT or BULLY player notices the opponent's mean or nice move immediately, in time for making the appropriate response in the same round. As noted earlier, this stacks the deck somewhat in favor of the reactive strategies -- and especially in favor of TIT FOR TAT whose opening nice gesture would otherwise leave it open to exploitation in the first round of play. The enormous advantage of the simplification is that it permits us to limit attention to single-round interactions[21] while still allowing for multi-generation (evolutionary) developments.

In Matrices 3 and 4, the underlying Prisoners' Dilemma and Chicken environments have been expanded by adding TIT FOR TAT to the available menus of strategies. The only cells needing explanation are at the lower right, the payoffs when TIT FOR TAT players encounter one another. Following the definition of TFT, each side makes a nice opening gesture to which the other player would respond in kind, so evidently this is a nice-nice behavioral interaction. Thus, the TFT-TFT payoffs are (3,3) in both games -- the same payoffs as COOPERATE-COOPERATE in Prisoners' Dilemma and COWARD-COWARD in Chicken.

Matrices 5 and 6 tabulate the corresponding payoffs when BULLY is the added reactive strategy, assuming instant response as before. Once again, the only problematic point is the outcome when two BULLY players encounter one another. Since BULLY involves an initial mean move, it is reasonable to assume that whichever party is in a position to move first (to take the offensive, so to speak) opens with a hostile gesture, in which case the other will back away -- i.e., will respond with nice. But which of the two players has the opportunity to go first will be a random matter. It follows that BULLY-BULLY will involve mean-vs.-nice interactions in which, on average, each player can expect to have the advantageous mean side of the encounter half the time and the disadvantageous nice side the rest of the time. Thus, in Matrices 5 and 6, the outcome of the BULLY-BULLY interaction in Prisoners' Dilemma is $.5(4+1) = 2.5$ for each player, while in Chicken the corresponding average payoffs are $.5(4+2) = 3$ for each.

2. THE MODEL -- BASIC ELEMENTS

Let a_{sj} be the payoff obtained in any round by an individual playing strategy s against strategy j. The mean return or yield Y_s to strategy s in that round will depend upon the population fractions p_j playing the various strategies:

$$(1) \qquad Y_s = \Sigma_j p_j a_{sj}$$

The corresponding overall mean yield Y for the population as a whole is:

$$(2) \qquad Y = \Sigma_s p_s Y_s$$

Let F_s denote the relative fitness of strategy s -- the difference between the yield of strategy s and the overall mean yield in the population:

$$(3) \qquad F_s = Y_s - Y$$

Of course, the weighted mean fitness for the population as a whole must be zero.

Under selection, the representation of strategy s in the population will change in proportion to its relative fitness F_s. In a discrete-generation dynamic model, the change in p_s will be denoted Δp_s. The parameter k reflects the rapidity of change or "sensitivity" of the process. Thus:

$$(4) \qquad \Delta p_s = kp_sF_s = kp_s(Y_s - Y)$$

Equation (4) satisfies the logical requirement that the sum of the increments Δp_s must be zero.[22,23]

3. EVOLUTIONARY EQUILIBRIUM CONCEPTS

Considerable confusion has been caused by the fact that (at least) five different concepts of evolutionary equilibrium have played some role in the literature.

1. Critical Point (CP): At any moment of time each of the S strategy options will be represented by a certain population proportion p_s. A necessary though by no means sufficient condition for an evolutionary equilibrium is that the vector $(p_1 \ldots p_S)$ is stationary with regard to the dynamic process (4). Any vector of population proportions meeting the stationarity condition is termed a Critical Point.[24]

Formally, the condition for a vector of strategies to be a Critical Point (CP) is:

(5) If $p_s > 0$, then $F_s = 0$ Condition for CP

That is, to have stationarity, every strategy with positive representation must have zero fitness (i.e., its yield must exactly equal the average population yield), implying no net tendency to increase or decrease in the next generation.[25]

There are three general categories of Critical Points: (i) at a vertex (all the population is following a single strategy), (ii) along an edge (more than one strategy, but not all of the strategies, have positive representation), and (iii) in the interior (all of the available strategies are represented). From equation (5), it follows as a corollary that: Every vertex is necessarily a Critical Point. (When only a single strategy s has positive representation p_s, its relative fitness F_s must be zero.) Points along edges or in the interior may or may not be CP's.

2. Evolutionary Equilibrium Point (EEP): The CP condition is evidently a

very weak one. Let us jump now to the condition for an Evolutionary Equilibrium Point (EEP), which is the strongest (and the correct) equilibrium concept for evolutionary dynamics.

Definition: An Evolutionary Equilibrium Point (EEP) is a Critical Point for which all the points in its neighborhood belong to its convergency zone.

A Critical Point may or may not have a convergency zone -- that is, a set of population distributions for which the evolutionary process (4) leads into that CP as terminus. For a CP to be an EEP, it must be that all the points sufficiently close to it fall within its convergency zone. This means that an EEP is immune to (sufficiently small) shocks: any momentary displacement would initiate a dynamic process restoring the previous situation. In contrast, a CP that has only a partial convergency zone, or no convergency zone at all, will not be stable under even small shocks.

3. Nash Equilibrium (NE): Standing logically somewhere between the CP and EEP concepts is a condition that Axelrod inappropriately terms "collective stability" (Axelrod [1984], p. 56). In his non-technical discussion, Axelrod always interprets any distribution satisfying this condition as an Evolutionary Equilibrium Point, that is, as a population distribution resistant to shocks ("invasions").[26] In his logical analysis, however, he defines a strategy s as "collectively stable" whenever no other strategy s′ has strictly higher payoff in an all-s population.[27] Consequently, the requirement for so-called "collective stability" is nothing but the familiar Nash Equilibrium (NE) condition.[28]

More formally, the Nash Equilibrium condition can be expressed:

(6) $$F_s(S) \geq F_j(S)$$

This means that, when all the population is following strategy s (as indicated by the capital letter S within parentheses), then s has no less fitness than any potential invader strategy j.

A strategy s can in principle be a probabilistic mixture in which strategies $s_1, s_2, \ldots$ are being played in the proportions $p_1, p_2, \ldots$. However, instead of thinking of a uniform population each member of whom is playing a mixed strategy, we can equivalently think in terms of a mixed population: each individual is playing a pure strategy, but the different pure strategies $s_1, s_2, \ldots$ are represented in the population in the proportions $p_1, p_2, \ldots$.[29] (NOTE: If strategy s is a mixture of two strategies s_1 and s_2, then what would be the s-vertex in a space including mixed strategies becomes a point along the s_1s_2-edge in a space of pure strategies only.) We will always be employing the latter geometry.

The NE condition is stronger (less inclusive) than the CP condition, since all vertices are CP's but not all of them will satisfy condition (6). Nevertheless, (6) is not strong enough to rule out successful invasion. To be stable with regard to shocks, a population vector must have a convergency zone that includes all the points in its neighborhood. But as our illustrations will show, a distribution satisfying the Nash Equilibrium condition may well have an incomplete convergency zone, or even no convergency zone whatsoever!

4. Evolutionarily Stable Strategy (ESS): The reason why the NE condition is not strong enough to preclude invasion will be clarified by the fourth equilibrium concept to be taken up, the Evolutionarily Stable Strategy (ESS) as defined by John Maynard Smith:[30]

$$F_s(S) > F_j(S) \quad \text{or} \quad F_s(S) = F_j(S) \text{ and } F_s(J) > F_j(J) \tag{7}$$

As in the NE condition (6), mixed as well as pure strategies s are allowed -- but once again, mixed strategies will be understood as referring to mixed populations.

Evidently, the ESS condition is not satisfied, as is the NE condition, by <u>all</u> the possible situations where $F_s(S) = F_j(S)$ but only by the subset of them indicated by the second condition on the lower line of (7). Thus, for an ESS, a strategy s must have <u>strictly</u> superior fitness when the population is all-s, <u>or</u>, if an equality with some other strategy j holds there, then s must be strictly superior when the population is all-j.[31]

Even Maynard Smith's ESS, it has been shown,[32] is not fully equivalent to the true EEP condition for evolutionary equilibrium. The reason, essentially, is that the ESS is defined solely in terms of the <u>payoffs</u>, whereas evolutionary convergence depends not only upon these but upon the dynamics. There are ESS's that cannot be attained under certain dynamics, and there are dynamics that may lead to EEP's that do not satisfy the conditions for an ESS. However, for most purposes the ESS is a satisfactory evolutionary equilibrium condition.

5. <u>Evolutionary Equilibrium Region</u>: One final equilibrium concept will also be relevant for the analysis that follows:

> <u>Definition</u>: An Evolutionary Equilibrium <u>Region</u> (EER) is a compact set of points, no one of which is an EEP, whose entire neighborhood belongs to the convergency zone of <u>some</u> point within the region.

If a set of points constitutes an EER, no single point has to have a convergency zone that covers its entire neighborhood. In other words, no single point is an "attractor" <u>but the region as a whole is</u>.[33]

Certain of the conditions for an EEP and an EER, for 3x3 interactions, are summarized in Table II.

4. ILLUSTRATIVE EVOLUTIONARY EQUILIBRIA FOR SIMPLE 2x2 INTERACTIONS

As background for the developments to follow, this section illustrates the characteristic evolutionary equilibria of the simple 2x2 Prisoners' Dilemma and Chicken games.

Using Matrix 1 for the numerical payoffs under Prisoners' Dilemma, Figure 1a is one possible "historical" picture of evolutionary progression over 500 generations -- starting from arbitrary initial proportions $(p_C\ p_D) = (.7\ .3)$. DEFECT seems to be driving COOPERATE to extinction, i.e., the terminus of the dynamic process appears to be $(p_C\ p_D) = (0\ 1)$. This is indeed what will occur, regardless of the initial position (so long as DEFECT is initially present in the population). The reason is indicated in Figure 1b, where we see that the DEFECT yield Y_D remains always greater than the COOPERATE yield Y_C until the equilibrium at $(p_C\ p_D) = (0\ 1)$ is reached. Such a clear result comes about when one strategy strictly dominates another in the payoff matrix. In this case, of course, the DEFECT payoffs strictly dominate the COOPERATE payoffs in Matrix 1.

In terms of the equilibrium concepts of the preceding section:

1. <u>Critical Point</u>: The C and D vertices are the only CP's. That is, the only stationary population distributions under the dynamic process (4) are all-COOPERATE and all-DEFECT.

2. <u>Evolutionary Equilibrium Point</u>: For the reason explained above, the D vertex (the all-DEFECT population) is the sole EEP.

3. <u>Nash Equilibrium</u>: As is well-known, the D vertex is also the sole Nash Equilibrium in Prisoners' Dilemma. (So, in this case, the NE and EEP criteria coincide.)

4. <u>Evolutionarily Stable Strategy</u>: The same result: only the D vertex satisfies condition (7). From Matrix 1 it is easy to verify that an all-

DEFECT population meets the strong condition in the upper line of equation (7).

5. Evolutionary Equilibrium Region: There is no EER (other than the degenerate region represented by the single EEP at the D vertex).

Turning to Chicken, Figure 2a is an historical diagram using the numerical payoffs of Matrix 2. Starting from the arbitrary initial distribution $(p_C\ p_D) = (.7\ .3)$, the population appears to be converging toward a mixture terminus at $(p_C\ p_D) = (.5\ .5)$. This follows of course from the way the yields Y_C and Y_D approach equality over time.[34]

Briefly summarizing the equilibrium conditions for Chicken:

1. Critical Point: The C and D vertices, as always, are CP's. But now the mixture terminus $(p_C\ p_D) = (.5\ .5)$ is also a CP. Such a population will be stationary under the dynamic process (4) since, at that mixture, the yields Y_C and Y_D are equal.

2. Evolutionary Equilibrium Point: As suggested by the illustration in Figures 2a and 2b, the mixed terminus is the sole EEP.

3. Nash Equilibrium: There is no NE in pure strategies.[35] However, there is an NE in mixed strategies, corresponding numerically to the mixed-population EEP at $(p_C\ p_D) = (.5\ .5)$.

4. Evolutionarily Stable Strategy: This mixed terminus is also an ESS. At the equilibrium $(p_C\ p_D) = (.5\ .5)$, however, the yields Y_C and Y_D are both equal to the yield to the equilibrium mixed strategy. It follows that the operative portion of the ESS condition is that indicated on the lower line of (7). As before, we will interpret the evolutionary outcome as a mixed population of individuals following pure strategies rather than as a uniform population playing a mixed strategy.

5. Evolutionary Equilibrium Region: There is no EER, other than the degenerate region represented by the single EEP at the mixed terminus.

In general, therefore, it appears that in the 2x2 Prisoners' Dilemma and Chicken games, the three solution concepts EEP, NE, and ESS coincide. However, this will not ordinarily be the case when we expand the menu of options to include the reactive strategies TIT FOR TAT and BULLY.

5. PRISONERS' DILEMMA -- TIT FOR TAT VS. BULLY AS REACTIVE STRATEGIES

In this section the basic 2x2 matrix for Prisoners' Dilemma (Matrix 1) is expanded to a 3x3 interaction by adjoining either TIT FOR TAT (Matrix 3) or else BULLY (Matrix 5) as reactive strategy. (The section following provides a corresponding analysis of Chicken.)

Starting with TIT FOR TAT as the third strategy, a typical "historical" development based upon the payoffs of Matrix 3 is pictured in Figure 3, where the initial proportions were set at $(p_D \quad p_C \quad p_T) = (1/3 \quad 1/3 \quad 1/3)$. As can be seen, DEFECT is being driven to extinction; however, the evolutionary terminus does not appear to be all-TIT FOR TAT but rather a mixture of TIT FOR TAT and COOPERATE.

This is confirmed, and the process illustrated with greater generality, in the "phase diagram" of Figure 4. Each point of the triangle represents a vector of population proportions $(p_D \quad p_C \quad p_T)$, where of course $\Sigma_s p_s = 1$. Specifically, at any point in the triangle the proportion p_D playing DEFECT is represented by the horizontal distance from the origin, the proportion p_C playing COOPERATE by the vertical distance from the origin, while the remaining proportion p_T is shown geometrically by the distance (measured either vertically or horizontally) from the point to the hypotenuse. Thus the lower-right vertex of the triangle represents an all-DEFECT population, the upper-left vertex is an all-COOPERATE population, and the origin is an all-TFT population.

In this and the other phase diagrams, 200 evolutionary trajectories were calculated starting from initial positions spread approximately evenly throughout the triangle. Every path has 15 points (dynamic iterations in accordance with equation (4)). The large open arrows indicate the general directions of the evolutionary trends. Regions where dynamic change is occurring at a relatively slow pace are indicated by relatively short trajectories in the picture, while darker clouds suggest regions that tend to accumulate trajectories. Since it is an attractor, we would expect to see such a cloud surrounding an Evolutionary Equilibrium Point (EEP). And since the region as a whole is an attractor, we expect to find a somewhat more diffuse cloud in the neighborhood of an Evolutionary Equilibrium Region (EER). However, as will be seen, depending upon the particulars of the situation there could be some concentration of trajectories in regions not associated with an EEP nor an EER.

In Figure 4 the curve LD, representing the set of points for which $F_D = 0$ -- that is, for which DEFECT has zero relative fitness in comparison with the rest of the population -- divides the triangle into two portions. The vertical coordinate of point L is .5, as plotted in the diagram.[36] Above LD, the relative fitness of DEFECT is positive: that is, $F_D > 0$ or equivalently $Y_D > Y$. Consequently, DEFECT will be increasing in the population, as suggested by the upper open arrow. However, all the trajectories eventually cross LD, so as to enter the portion of the triangle where the relative fitness of DEFECT is negative. In this region below LD, the large open arrow indicates that DEFECT will be decreasing. Thus, ultimately, DEFECT goes extinct.[37]

Summarizing in terms of the different equilibrium concepts of the preceding section:

Critical Points: As always the vertices are CP's. To have an interior Critical Point, it follows immediately from equations (4) and (5) that all three $F_s=0$ curves must intersect. But it is easily seen that the $F_C=0$ and the $F_T=0$ curves both overlie the vertical axis, so there can be no interior Critical Point. (This conclusion also follows directly from the condition in Part C of Table 2, since Matrix 3 indicates that it is not the case that each strategy is strictly inferior at its own vertex to some other. In fact, no one of the three strategies is strictly inferior at its own vertex.) As for possible edge CP's, along the left edge TC of the diagram, as previously noted, the relative fitnesses F_C and F_T for COOPERATE and TIT FOR TAT are both zero. Since DEFECT is unrepresented anywhere along the vertical axis, it follows that each and every point in the portion TL of the vertical edge satisfies the condition for a Critical Point.

Nash Equilibria: From Matrix 3 of Table 1, it is evident that the DEFECT vertex and the TIT FOR TAT vertex -- but not the COOPERATE vertex -- are indeed NE's. In addition, each and every mixture of DEFECT and COOPERATE represented by points along TL is also an NE.

Evolutionarily Stable Strategies/Evolutionary Equilibrium Point: However, as Figure 4 suggests, none of these NE's are EEP's. The DEFECT vertex has no convergency zone whatsoever.[38] And the TIT FOR TAT vertex, together with each and every mixture of TIT FOR TAT with COOPERATE that satisfies the NE condition, has a convergency zone of measure zero, consisting only of a single curve in the diagram. In fact, there is no ESS or EEP here.

Evolutionary Equilibrium Region: However, it follows immediately from the discussion above, as is also evident in the diagram, that the TL range as a whole (with the exception of the singular point L) satisfies the condition given above for an Evolutionary Equilibrium Region (EER). In other words, any point in a sufficiently near neighborhood of the semi-closed interval TL is attracted to some point within the region.

The evolutionary equilibrium in the expanded Prisoners' Dilemma game, TIT FOR TAT being the added (reactive) strategy, is therefore a range of population mixtures of COOPERATE and TFT. Since this is an edge solution rather than a vertex solution proper, to some extent the objection we raised earlier -- that models with a vertex solution in which only a single strategy survives in equilibrium are unsatisfying -- is met. Still, the two surviving strategies are both "nice," whereas in the real world we really expect to find both "nice" and "mean" behaviors maintaining themselves in the population. To explore this possibility, we now turn to BULLY as an alternative reactive strategy.

In Figure 5, the origin now represents the all-BULLY vertex, so it is the proportion of the population playing BULLY (p_B) that is scaled by distance from the hypotenuse of the triangle. As can be seen, the cloud of points in Figure 5 is dense only in the neighborhood of the all-DEFECT vertex, which indeed is the sole attractor. Listing the different equilibrium concepts as before, we now have:

Critical Points: The three vertices are the only CP's.

Nash Equilibria: The all-DEFECT vertex is the sole NE.

Evolutionarily Stable Strategies/Evolutionary Equilibrium Point: The all-DEFECT vertex is also the sole ESS and EEP.

Evolutionary Equilibrium Region: There is no EER, apart from the degenerate region represented by the single EEP.

Thus, BULLY as reactive strategy has no success at all. That BULLY does so poorly is not surprising when we notice that its payoffs in Matrix 6 are strictly dominated by DEFECT.[39]

Summing up with regard to evolutionary equilibrium conditions for the 3x3 Prisoners' Dilemma: (i) If TIT FOR TAT is the reactive strategy, no Evolutionary Equilibrium Point (EEP) exists. However, there is an Evolutionary Equilibrium Region (EER) consisting of a range of mixtures of TIT FOR TAT and COOPERATE in which the latter is represented in sufficiently small proportion. (ii) With BULLY as reactive strategy, in contrast, all-DEFECT is the sole EEP.

The distributions of populations over time as reflected in the probability clouds are also of interest. In Figure 4, where TIT FOR TAT is the reactive strategy, while there is a certain concentration in the neighborhood of the EER (the range TL along the vertical axis), the upper portion of this range is rather bare. The reason is that the initiating points of the trajectories were spread approximately uniformly over the entire triangle. As can be seen in the diagram, the lower portion of TL is a terminus that draws from a much larger area of possible starting positions than does the upper portion. When BULLY is the reactive strategy, Figure 5 indicates only a single cloud in the neighborhood of the single EEP at the all-DEFECT vertex.

6. CHICKEN -- TIT FOR TAT VS. BULLY AS REACTIVE STRATEGIES

The analysis of the Chicken environment can proceed in more summary fashion. We can concentrate attention upon the true equilibrium conditions: presence or absence of an Evolutionary Equilibrium Point (EEP) or Evolutionary Equilibrium Region (EER).

With TIT FOR TAT serving as the added reactive strategy (Matrix 4), Figure 6 is the analog of the corresponding Figure 4 for Prisoners' Dilemma. The curve QD divides the diagram into two regions, in a way somewhat analogous to LD in Figure 4. However, QD here is a separatrix. That is, QD divides the "basins" in which trajectories lead to different termini.

Without extensive discussion of the details, the open arrows suggest an EEP at point G along the CD edge and possibly an EER along a portion of the vertical axis as in Figure 5. And in fact, both of these suggestions are correct. There are two classes of solutions:

(i) For initial population distributions lying above the separatrix QD, the trajectories all converge to an EEP at the distribution $(p_C \;\; p_D \;\; p_T) = (.5 \;\; .5 \;\; 0)$. As this point is along the hypotenuse of the triangle, TIT FOR TAT has gone extinct!

(ii) For initial distributions lying below the curve QD, an EER is approached along the vertical axis, from the T vertex up to point R representing the mixture $(p_C \;\; p_D \;\; p_T) = (2/3 \;\; 0 \;\; 1/3)$.

In Figure 7, BULLY replaces TIT FOR TAT as the reactive strategy. Here there is an EEP along the horizontal axis at the population distribution $(p_C \;\; p_D \;\; p_B) = (0 \;\; .5 \;\; .5)$ -- that is, an equal mixture of DEFECT and BULLY, with COOPERATE having been driven out of the population. (There is also a CP midway along the hypotenuse, at point N, but this is not an EEP.)

Summing up for the Chicken payoff environment:

(i) With TIT FOR TAT as reactive strategy, just as in Prisoners' Dilemma there is an Evolutionary Equilibrium Region (EER) involving cooperative behavior in the form of a mixture of COWARD and TIT FOR TAT. But, in contrast with the results in the Prisoners' Dilemma environment, from certain initial positions the dynamic process may lead instead to an Evolutionary Equilibrium Point (EEP) excluding TIT FOR TAT in favor of a mixture of COWARD and DAREDEVIL. This latter outcome suggests the "dominance hierarchies" observed in some biological interactions. Such a solution becomes viable in the Chicken environment because it is better to be the exploited victim of a nice-mean pair (payoff 2 in Matrix 5) than to be involved in an interaction with hostile behaviors on both sides (payoff 1).

(ii) With BULLY as reactive strategy, the only equilibrium is an EEP involving a mixture of DAREDEVIL and BULLY. Thus, BULLY is viable in the Chicken environment, but its presence does not lead the population toward more cooperative behavior patterns.

7. 4x4 INTERACTIONS

This section examines what happens when both of the reactive strategies TIT FOR TAT and BULLY are in play -- together of course with the archetype strategies defining the Prisoners' Dilemma or the Chicken environment.

In Table III, Matrix 7 shows the basic 2x2 payoffs for a Prisoners' Dilemma environment, while Matrix 8 indicates the effect of simultaneously adding TIT FOR TAT and BULLY as strategy options. For a reason to be explained below, in Matrix 7 the underlying 2x2 numerical payoffs for the archetype COOPERATE and DEFECT strategies differ slightly from those in Matrix 1. In fact, the 2x2 archetypical payoffs of Matrix 7 here are the same as those used by Axelrod in his famous tournaments.

The evolutionary dynamics corresponding to the 4x4 interaction of Matrix 8 are shown in the "unfolded tetrahedron" of Figure 8.[40] This tetrahedron is to be imagined as standing on the triangle T-D-B as base, with the C vertex as the tip. (Upon folding the tetrahedron into its three-dimensional form, all the C vertices will come together.) Only the dynamic movements along the four faces are pictured; we cannot directly see what is happening in the interior, though to some extent that can be inferred from developments along the various faces.

How the numerical payoffs in Matrix 8 are derived from the underlying 2x2 elements of Matrix 7 will mainly be evident from what has gone before. The only novelty is the encounter between the two reactive strategies -- TIT FOR TAT and BULLY. For this interaction, the payoffs were computed on the assumptions that: (1) In each round, a reactive player will be able to detect and instantly respond (in TIT FOR TAT or BULLY fashion, depending upon his own type) to the opponent's opening move, but will not discover (or at any rate will not respond to) the opponent's underlying strategy. (2) There will be an indefinite number of rounds of play. On these assumptions, there are two possible sequences of events, depending upon who makes the opening move:

(i) When the TIT FOR TAT player has the first move, he will of course open with "nice." And, following his nature in turn, BULLY will respond with "mean." So, in the first round, the TFT,BULLY payoffs are (0,5). In the second round, TFT will mirror BULLY's previous "mean" by opening with "mean" himself, to which BULLY will respond with "nice" -- making the second-round payoffs (5,0). This alternation will keep up forever, making the long-run average payoff 2.5 for each side.

(ii) However, with equal probability, BULLY moves first. If so, his "mean" is answered with TFT's "mean," so the first-round payoffs are (1,1). Then in the second round, BULLY will respond to TFT's previous "mean" with "nice," so TFT will also play "nice," making the second-round payoffs (3,3). Again, this alternation will keep up forever, the long-run average being 2 each.

Finally, since these two sequences are equally probable, the payoffs shown in Table 8 are the expectations -- 2.25 to each side.

Turning now to the evolutionary dynamics of Figure 8, the 3x3 trajectories shown on the C-D-T and C-D-B faces of the tetrahedron are entirely consistent with the results described in Section 5 above. Specifically, the C-D-T interaction once again has no Evolutionary Equilibrium Point (EEP) but does have an Evolutionary Equilibrium Region (EER) -- which starts at the T vertex and extends some partial distance toward the C vertex along the TC edge. And the C-D-B interaction as before has only a single EEP at the B vertex. The C-T-B and D-T-B triangles, however, represent strategy triads not previously considered; on each of these faces, the two reactive strategies are matched up together against only one of the archetype strategies --

COOPERATE or BULLY. Looking first at the triangle D-T-B (the base of the tetrahedron), there is a single EEP at the T vertex. That is, when TIT FOR TAT and BULLY are both matched against DEFECT, only TFT survives. On the other hand, triangle C-T-B has both an EEP at vertex B and an EER along CT.

Examination of the faces of the tetrahedron therefore suggests a number of possible equilibria. For each such candidate, the question is whether or not a complete convergency zone -- now including trajectories passing through the interior -- exists. These interior trajectories are all repelled from the C vertex, so as to be driven toward the D-T-B base of the tetrahedron. It turns out that the only equilibrium for this 4x4 interaction is the same kind of EER as obtained in the initial COOPERATE-DEFECT-TFT encounter: that is, an attractor region starting at the TIT FOR TAT vertex and moving some partial distance along the TC edge toward the COOPERATE vertex. (The rather dark probability cloud near the D vertex suggests a possible EEP there, but this turns out not to be the case; all the trajectories in that neighborhood eventually diverge toward the EER.) Once again an all-TIT FOR TAT population has only vanishingly small probability of occurring, but the surviving strategies will both be "nice."

In setting up his tournaments, Axelrod was careful to select numerical payoffs for the underlying 2x2 Prisoners' Dilemma such that the reward to each party in a COOPERATE-COOPERATE encounter (3 and 3 in Matrix 7) exceeded the average of the asymmetrical payoffs when COOPERATE meets DEFECT (5 and 0 in Matrix 7).[41] The purpose was to make it unattractive for a pair of players to alternate roles in a complementary COOPERATE-versus-DEFECT pattern. But, we have seen, just such an alternation tends to come about without deliberate design when the reactive strategies TIT FOR TAT and BULLY encounter one another. So it will be of interest to consider a payoff matrix tending to favor rather than disfavor such a pattern. Matrix 9 is an example of such "asymmetry-favoring" 2x2 payoffs, leading in turn to the expanded 4x4 tabulation shown in Matrix 10.

The unfolded tetrahedron in Figure 9 pictures the evolutionary implications of Matrix 10. There are a number of interesting details, including (i) strong convergence toward B in the C-T-B triangle; (ii) the outward-spiralling "hurricane" in the interior of the D-T-B triangle, and (iii) the dark probability cloud on all the faces in the neighborhood of the D vertex. But proceeding directly to the question of equilibrium, it turns out that there is none here! The attractor for all trajectories is a cycle T -> B -> D -> T along the edges of the base of the tetrahedron. Each of the vertices is of course a Critical Point (CP), but none has a complete convergency zone. Thus, COOPERATE is rapidly driven out, but over time the three other strategies -- TIT FOR TAT, BULLY, and DEFECT -- will successively predominate in the population in a never-ending cycle.

For the Chicken payoff environment, Matrices 11 and 12 are the 2x2 and 4x4 patterns under a "symmetry-favoring" assumption, leading to the picture in Figure 10. Skipping over the details, we see once again the familiar type of EER along the TC edge, starting at the T vertex and moving some distance toward the C vertex. And, there is a single EEP along the BD edge. Thus, while the types of equilibria differ, in terms of behaviors there are two qualitative sorts of outcomes in Chicken: a "nice" mixture of TIT FOR TAT and COOPERATE, and a "mean" mixture of DEFECT and BULLY.

Finally, Matrices 13 and 14 represent an "asymmetry-favoring" payoff assumption for Chicken. The corresponding Figure 10 now has only a single equilibrium, an EEP along the BD edge. The change in the payoff assumption has eliminated the nice-behavior EER along the TC edge.

8. SUMMARY AND DISCUSSION

1. Axelrod and his followers have made strong claims for the success of TIT FOR TAT in evolutionary competitions. Our previous paper showed that these results were not very robust, holding only under a narrow window of assumptions. Quite different outcomes were generated, among other ways, by postulating an elimination contest rather than a round-robin tournament, by providing for costs of complexity, or by allowing some risk of incorrect identification of the opponent's type. In addition, we showed, even under Axelrod's assumptions the evolutionary outcome in Prisoners' Dilemma is not all-TIT FOR TAT but rather a mixture of TIT FOR TAT and COOPERATE.

2. This paper compares the evolutionary success of TIT FOR TAT and the alternative reactive strategy BULLY, in both the Prisoners' Dilemma and Chicken payoff environments. Where TIT FOR TAT involves an opening "nice" move and an echoing response thereafter, BULLY is defined by an opening "mean" move and a reversing response thereafter. In Prisoners' Dilemma, the nice and mean archetype strategies are COOPERATE versus DEFECT; in Chicken, we call them COWARD and DAREDEVIL. We considered both the 3x3 interactions when either TIT FOR TAT or BULLY is in play against the archetype strategies of Prisoners' Dilemma and Chicken, and the 4x4 interactions when both of the reactive strategies are simultaneously in play.

3. Axelrod's so-called "collective stability" concept was shown to be equivalent to the familiar Nash Equilibrium (NE) condition. NE is not strong enough for evolutionary equilibrium, as it counts as an equilibrium any point for which a possible invader is not strongly superior in fitness. We showed that successful invasion may be possible even if the invader has only equal fitness. A true Evolutionary Equilibrium Point (EEP), which is for our purposes equivalent to Maynard Smith's condition for an Evolutionar-

ily Stable Strategy (ESS), must be an "attractor": every population distribution in its neighborhood must lie in its convergency zone. Also, we defined a set of points as an Evolutionary Equilibrium Region (EER) when, even though no individual point is an EEP, every distribution in the neighborhood of the region as a whole is attracted to some point within the region.

4. Substantively, the main results for the 3x3 interactions are: Prisoners' Dilemma: With TIT FOR TAT as reactive strategy there is no single Evolutionary Equilibrium Point (EEP). Instead, the outcome is an Evolutionary Equilibrium Region (EER) containing an admixture of free-riding COOPERATE players with TIT FOR TAT. (But note that in terms of behavior, only "nice" play survives.) With BULLY as the reactive strategy, on the other hand, the all-DEFECT EEP is the sole evolutionary equilibrium.

Chicken: With TIT FOR TAT as reactive strategy, there are two equilibria serving as attractors for trajectories in two different "basins" of the phase diagram. One of these is analogous to the Prisoners' Dilemma outcome: a mixture of TIT FOR TAT with the other "nice" strategy -- COWARD. The other attractor is an EEP representing a mixture of COWARD and DAREDEVIL, where TIT FOR TAT has gone extinct. With BULLY as reactive strategy, the only equilibrium is an EEP representing a mixture of BULLY and DAREDEVIL; here only "mean" behavior can survive.

5. The 4x4 interactions, where TIT FOR TAT and BULLY are both in play in addition to the archetype strategies, displayed a number of interesting patterns. Under the "symmetry-favoring" payoffs postulated by Axelrod, in Prisoners' Dilemma the evolutionary result obtained for the 3x3 case with TIT FOR TAT as the only reactive strategy -- an EER representing a mixture of TIT FOR TAT together with a certain proportion of the other "nice" strategy -- held also for the 4x4 case. This was true also for Chicken,

where, however, there was once again an EEP excluding the nice strategies and involving a mixture of DAREDEVIL and BULLY. "Asymmetry-favoring" payoffs, in contrast, entirely eliminated the "nice" EER. Instead, for Prisoners' Dilemma the evolutionary progression approached a limit cycle where the population is predominantly of a single type but that type follows the time-path TIT FOR TAT -> Bully -> DEFECT -> TIT FOR TAT... As for Chicken, under asymmetry-favoring payoffs the sole equilibrium is an EEP representing a mixture of the mean strategies BULLY and DAREDEVIL.

6. Summarizing, in support of Axelrod's claims it appears there are a number of distinct situations which, even if they do not lead to an all-TIT FOR TAT population, at least imply the survival only of nice strategies in evolutionary equilibrium. But, as against these claims, even holding in the main to all his other assumptions, there are many situations leading to the total or partial extinction of TIT FOR TAT or, more generally, of nice strategies.

7. Our diagrams revealed a variety of dense probability clouds, population distributions that tend to accumulate trajectories even though not perhaps ultimately viable as equilibrium points. Since environments rarely remain stable enough for final evolutionary equilibrium to be achieved, the existence and location of such probability clouds may be of predictive significance.

8. Methodologically, for consistency with what is observed in the real world, one would like to have models explaining the simultaneous survival of nice and mean behaviors in populations. <u>None</u> of the interactions studied here led to such a result! In some cases only nice strategies survived, in others only mean ones. There were also situations with multiple possible evolutionary results (where different basins led to different termini), but

each possible outcome was either all-nice or all-mean. (And, there was one case with an equilibrium cycle over mean and nice behaviors.) Our previous paper[42] showed that a stable mixture of mean and nice behaviors can indeed come about, provided that the payoff matrix is modified to permit a PUNISHER strategy. That is, if there are "bounty hunters" who can actually profit at the expense of DEFECT or DAREDEVIL, as opposed to TIT FOR TAT and BULLY players who at best only hold their own against them. The rather unsatisfying results here, as to the existence of evolutionary equilibria displaying a mixture of mean and nice behaviors when only TIT FOR TAT and BULLY are available as reactive strategies, suggest that PUNISHER-type strategies warrant more attention than they have as yet received.

ENDNOTES

[1]Valuable assistance was provided by Jorge Palamara. For helpful discussions we thank David Hirshleifer, Eric A. Smith, Ulrich Mueller, John Nachbar, and Heinrich Ursprung.

[2]As summarized in Axelrod [1984]. Axelrod and Dion [1988] is an updated survey.

[3]"The framework is broad enough to encompass not only people but also nations and bacteria." -- Axelrod [1984], p. 18. Such a claim has been pressed (perhaps beyond Axelrod's original intention) by enthusiastic reviewers and commentators such as Hofstadter [1985], Ch. 29.

[4]See Maynard Smith [1976, 1978], Maynard Smith and Price [1973], Hirshleifer [1982, pp. 13-38].

[5]The analytical foundations for his asserted conclusions are set forth in Axelrod [1984], Chapter 9 ("The Robustness of Reciprocity") and Appendix B ("Proofs of the Theoretical Propositions").

[6]Hirshleifer and Martinez Coll [1988], and see also the comment by Rapoport [1988]. Other significant critiques include Boyd and Leberbaum [1987] and Mueller [1988].

[7]At various points in his book Axelrod does caution the reader that not all social situations can be interpreted as Prisoners' Dilemmas. However, the main emphasis is the other way:

> "The advice given in this book to players of the Prisoner's Dilemma might also serve as good advice to national leaders as well.... Likewise, the

techniques discussed in chapter 7 for promoting cooperation in the Prisoner's Dilemma might also be useful in promoting cooperation in international politics." -- Axelrod [1984], pp. 190-191.

Even within the narrow class of two-party interactions where each player has exactly two behavioral options and the payoffs can be strictly ranked, Prisoners' Dilemma is just one of 78 qualitatively distinct payoff matrices -- see Rapoport and Guyer [1966]. Lipman [1986] has extended Axelrod's analysis to a somewhat wider class of environments.

[8]"The Soviet Union has occasionally taken steps which appear to be designed to prove the limits of its agreements with the United States. The sooner the United States detects and responds to these Soviet probes, the better." -- Axelrod [1984], p. 185. The recommendation is very likely sound, but the ongoing SU/US struggle can hardly be taken as an instance of <u>evolutionary</u> competition.

[9]Hirshleifer and Martinez Coll [1988], pp. 390-394.

[10]Hirshleifer and Martinez Coll [1988], p. 382.

[11]Hirshleifer and Martinez Coll [1988], p. 383. Mueller [1988] also examines the implications of mistakes in the form of "noise." In considering misperceptions as a possible factor impairing the effectiveness of TIT FOR TAT, Axelrod was surprised to find that TIT FOR TAT remained the best decision rule in his simulations. (Axelrod [1984], pp. 182-183, and also Axelrod and Dion [1988], p. 1387.) Our result was much stronger: we showed that TIT FOR TAT does notably <u>better</u> when it sometimes errs in identifying the opponent's type.

[12] This was also pointed out in Boyd and Lorberbaum [1987].

[13] Hirshleifer and Martinez Coll [1988], pp. 376-381. This result had already been obtained in Hirshleifer [1982], and see also Mueller [1988].

[14] The instant-response interpretation of TIT FOR TAT in the Prisoners' Dilemma context corresponds to the RETALIATOR strategy defined in Maynard Smith [1976 and 1982, Ch. 2] for the Hawk-Dove (i.e,, Chicken) environment. RETALIATOR is further analyzed, in terms of evolutionary equilibrium properties in both Prisoners' Dilemma and Chicken environments, in Hirshleifer [1982], pp. 24-25.

[15] Put another way, our assumption assures that the "shadow of the future" (Axelrod [1984], p. 126) is actually present in the here and now.

[16] See, for example, Maynard Smith and Price [1973], Zeeman [1981], Schuster and Sigmund [1985].

[17] The dynamic behavior of evolutionary systems away from equilibrium is also examined in Nachbar [1989].

[18] As definitions of Prisoners' Dilemma and Chick, the numbers in Matrices 1 and 2 are given only an ordinal interpretation. However, it will be convenient initially to interpret them as actual cardinal payoffs. As will be seen below, in some situations the actual cardinal magnitudes do make a difference for the qualitative nature of the results.

[19] TIT FOR TAT, since it is an imitative or echoing strategy, should logically have been called TIT FOR TIT! It is the reversing or contrarian strategy we call BULLY that ought to have been given the name TIT FOR TAT.

But reforming this flawed terminology is out of the question now.

[20]Properly speaking, there really are <u>four</u> rather than two simple reactive strategies, since elements of two dichotomies are involved: (i) whether the opening move is "nice" or "mean," and (ii) whether the response is to echo or to reverse the opponent's prior move. TIT FOR TAT combines a nice opening with an echoing response, while BULLY represents a nasty opening with a reversing response. The other two possible patterns are not without interest, but unfortunately cannot be analyzed here.

[21]In dealing with 4x4 interactions later on, however, we will be giving a multi-round <u>interpretation</u> to justify some of the numerical payoffs.

[22]Essentially the same dynamic equation is employed in the evolutionary models of Taylor and Jonker [1978], Zeeman [1981], Maynard Smith [1982], and Schuster and Sigmund [1985]. There are, however, other logically possible dynamics: see, for example, Friedman [1989].

[23]In using equation (4) to trace the evolution of the population proportions over time, there is a possibility that the indicated discrete change Δp_s in any given period will carry p_s out of its allowable range between zero and unity. To avoid this difficulty, we were careful in our simulations to use a sufficiently small k. Another way out, not adopted here, would be to employ a continuous dynamic equation, replacing Δp_s on the left-hand-side of (4) with the time-derivative dp_s/dt.

[24]The terminology here follows Abraham and Shaw [1983].

[25]If some strategy s is <u>not</u> represented, its fitness does not matter -- since, by equation (4), its p_s will remain unchanged anyway.

[26]"A strategy is collectively stable if no strategy can invade it." -- Axelrod [1984], p. 56.

[27]"To say that ALL D cannot invade TIT FOR TAT means that $V(\text{ALL D}|\text{TFT}) \leq V(\text{TFT}|\text{TFT})$." -- Axelrod [1984], p. 208.

[28]Lipman [1986] notices that Axelrod's so-called "collective stability" is equivalent to Nash Equilibrium (p. 325), but does not appear to appreciate that an NE point is not in general evolutionarily stable.

[29]Lipman [1986] provides such an extension of Axelrod's conditions.

[30]Maynard Smith [1976].

[31]In a footnote (Axelrod [1984], p. 217), Axelrod indicates that he is aware that his so-called "collective stability" condition does not quite correspond to the ESS or "evolutionary stability" condition of Maynard Smith. But he goes on to claim that essentially all his results would continue to hold under the ESS definition. This claim is refuted, among other ways, by our counterexamples showing that an all-TFT FOR TAT population is not stable in the evolutionary sense.

[32] Taylor and Jonker [1978], Friedman [1989].

[33]This condition corresponds to the "limit set" defined in Hirsch and Smale [1974], p. 240.

[34]However, the specific location of the mixture terminus will not in general be (.5 .5); it will depend upon the actual numerical values in the Chicken payoff matrix.

[35] We rule out here the asymmetrical (off-diagonal) Nash Equilibria -- where, in single-play encounters, one side chooses COWARD and the other DAREDEVIL. In the evolutionary process postulated, it is impossible for all encounters to be C versus D. A COWARD player cannot selectively arrange to encounter only DAREDEVILS. Instead, he will be randomly encountering both C and D players in proportion to their representation in the population.

[36] The specific location of point L along the vertical axis will depend, in general, upon the cardinal magnitudes in the payoff matrix.

[37] Here is where the distinction between the instant-response and the deferred-response assumption for TIT FOR TAT does make a difference. Axelrod asserted [1984, p. 211] that all-DEFECT is "always collectively stable" i.e., that all-DEFECT is always a Nash Equilibrium (NE). It is true that, as against instant-response TIT FOR TAT, all-DEFECT satisfies the NE condition -- but nevertheless, Figure 4 shows, an all-DEFECT population is not an Evolutionary Equilibrium Point (EEP). As against deferred-response TIT FOR TAT, however, all-DEFECT would be both an NE and an EEP. That is, for deferred-response TFT, a diagram analogous to Figure 4 would show a "basin" in which all the trajectories in its neighborhood lead toward the D vertex.

[38] As indicated in the preceding footnote, this assertion is valid when the instant-response definition of TIT FOR TAT holds. On the deferred-response assumption, however, the D vertex would have a complete convergency zone in its neighborhood, and so would be an EEP.

[39]A dominated strategy may sometimes survive as some positive proportion of the population at a Critical Point or at points within an Evolutionary Equilibrium Region, although not at an Evolutionary Equilibrium Point. For example, COOPERATE (which is dominated by TIT FOR TAT in Matrix 3), comprises a fraction of the population in the Evolutionary Equilibrium Region of Figure 1. The explanation is that, once in that region in Figure 1, COOPERATE has the same payoff as TIT FOR TAT. However, in Figure 4 DEFECT increasingly dominates BULLY as the all-DEFECT vertex is approached.

[40]Compare Maynard Smith [1982], p. 190.

[41]Axelrod [1984], pp. 8, 206-207. Our Matrix 1 above also satisfied this condition.

[42]Hirshleifer and Martinez Coll [1988], pp. 387-389.

References

Abraham, R.H. and Shaw, C.D., Dynamics, the Geometry of Behavior (Santa Cruz, Ca.: Aerial Press, Inc., 1983).

Axelrod, Robert, The Evolution of Cooperation (New York: Basic Books, 1984).

Axelrod, Robert and Douglas Dion, "The further evolution of cooperation," Science, v. 242 (9 December 1988), pp. 1385-1390.

Boyd, Robert and Jeffrey P. Lorberbaum, "No pure strategy is evolutionarily stable in the repeated Prisoner's Dilemma game," Nature, v. 327 (7 May 1987), pp. 58-59.

Friedman, Daniel, "Evolutionary Games in Economics," Economics Board, University of California, Santa Cruz, 1989.

Hirsch, M. and S. Smale, Differential Equations, Dynamical Systems and Linear Algebra (New York: Academic Press, 1974).

Hirshleifer, Jack, "Evolutionary models in economics and law: Cooperation versus conflict strategies," Research in Law and Economics, v. 4 (1982), pp. 1-60.

Hirshleifer, Jack and Juan Carlos Martinez Coll, "What strategies can support the evolutionary emergence of cooperation?" Journal of Conflict Resolution, v. 32 (June 1988), pp. 367-38.

Hofstadter, Douglas R., Metamagical Themas: Questing for the Essence of Mind and Pattern (New York: Basic Books, Inc., 1985).

Lipman, Barton L., "Cooperation Among egoists in Prisoners' Dilemma and Chicken games," Public Choice, v. 51 (1986), pp.315-331.

Maynard Smith, J., "Evolution and the theory of games," Amer. Scientist, v. 64 (Jan.-Feb. 1976).

Maynard Smith, J., "The evolution of behavior," Scientific American, v. 239 (Sept. 1978).

Maynard Smith, J., Evolution and the Theory of Games (Cambridge: Cambridge University Press, 1982).

Mueller, Ulrich, "Optimal retaliation for optimal cooperation," Journal of Conflict Resolution, v. 31 (Dec. 1987), pp. 692-724.

Nachbar, John H., "The Evolution of Cooperation in the Finitely Repeated Prisoner's Dilemma," The RAND Corporation, P-7591 (September 1989).

Rapoport, Anatol, "Editorial Comments on the article by Hirshleifer and Martinez Coll," Journal of Conflict Resolution, v. 32 (June 1988), pp. 399-401.

Rapoport, Anatol and Melvin Guyer, "A taxonomy of 2x2 games," General Systems, v. 11 (1966).

Schuster, P. and Sigmund, K., "Towards a dynamics of social behaviour: Strategic and genetic models for the evolution of animal conflicts," J. Social Biol. Struct., v. 8 (1985), pp. 255-277.

Taylor, P.D. and Jonker, L.B., "Evolutionarily stable strategies and game dynamics," Mathematical Biosciences, v. 40 (1978), pp. 145-156.

Zeeman, E.C., "Dynamics of the evolution of animal conflicts," Journal of Theoretical Biology, v. 89 (1981), pp. 249-270.

TABLE 1

Prisoners' Dilemma and Chicken --

Basic and Expanded Matrices

PRISONERS' DILEMMA

Matrix 1

	C	D
C	3,3	1,4
D	4,1	2,2

Matrix 3

	C	D	T
C	3,3	1,4	3,3
D	4,1	2,2	2,2
T	3,3	2,2	3,3

Matrix 5

	C	D	B
C	3,3	1,4	1,4
D	4,1	2,2	4,1
D	4,1	1,4	2.5,2.5

CHICKEN

Matrix 2

	C	D
C	3,3	2,4
D	4,2	1,1

Matrix 4

	C	D	T
C	3,3	2,4	3,3
D	4,2	1,1	1,1
T	3,3	1,1	3,3

Matrix 6

	C	D	B
C	3,3	2,4	2,4
D	4,2	1,1	4,2
B	4,2	2,4	3,3

TABLE 2

Conditions for EEP and EER in 3x3 Interactions

Notation: I: $i > j,k$ is shorthand for: "At vertex I (that is, where $p_i = 1$) the payoff to an i-player strictly exceeds the payoffs to either a j-player or a k-player." Equivalently, matrix payoff element a_{ii} exceeds a_{ji} and a_{ki}.

A. At vertex I

Necessary condition for EEP: I: $i \geq j,k$

Sufficient condition 1 for EEP: I: $i > j,k$

Explanation: If strategy i defeats both the other strategies at its own I vertex, then vertex I cannot be invaded and so will surely be an EEP.

Sufficient condition 2 for EEP: I: $i = j > k$ and J: $i > j$

Explanation: If at its own vertex I, strategy i defeats k but only ties strategy j, vertex I will still be an EEP provided that i defeats j at the latter's own vertex J.

B. Along edge IJ

Sufficient condition for EEP: I: $j > i > k$ and J: $i > j > k$

Explanation: If j defeats i at the I vertex, and i defeats j at the J vertex, and if the third strategy k is defeated by both of them at I and at J, there will surely be an EEP along the IJ edge.

Sufficient condition for EER: I: $i = j > k$ and J: $i = j$

Table 2 (cont.)

Explanation: As specified by this condition, the EER will consist of a portion of the IJ edge beginning at the I vertex. It will be the entire edge if J: $i = j \geq k$.

C. In the interior

Necessary condition: I: j or k > i and J: i or k > j and K: i or j > k

Explanation: As a necessary condition for an interior EEP, each strategy at its own vertex must be defeated by at least one of the other strategies.

Sufficient condition: I: j and k > i and J: i and k > j and K: i and j > k

Explanation: If each strategy does worst of all at its own vertex, there will surely be an interior EEP.

TABLE 3

Symmetry-Favoring Versus Asymmetry-Favoring Payoffs

2×2 PAYOFFS

Matrix 7

PRISONERS' DILEMMA

	C	D
C	3,3	0,5
D	5,0	1,1

Matrix 8

CHICKEN

	C	D
C	3,3	1,4
D	4,1	0,0

4×4 PAYOFFS

PRISONERS' DILEMMA

Symmetry-favoring

Matrix 9

	C	D	T	B
C	3,3	0,5	3,3	0,5
D	5,0	1,1	1,1	5,0
T	3,3	1,1	3,3	2.25,2.25
B	5,0	0,5	2.25,2.25	2.5,2.5

Asymmetry-favoring

Matrix 10

	C	D	T	B
C	2,2	0,7	2,2	0,7
D	7,0	1,1	1,1	7,0
T	2,2	1,1	2,2	2.5,2.5
B	7,0	0,7	2.5,2.5	3.5,3.5

CHICKEN

Matrix 11

	C	D	T	B
C	3,3	1,4	3,3	1,4
D	4,1	0,0	0,0	4,1
T	3,3	0,0	3,3	2,2
B	4,1	1,4	2,2	2.5,2.5

Matrix 12

	C	D	T	B
C	2,2	1,6	2,2	1,6
D	6,1	0,0	0,0	6,1
T	2,2	0,0	2,2	2.25,2.25
B	6,1	1,6	2.25,2.25	3.5,3.5

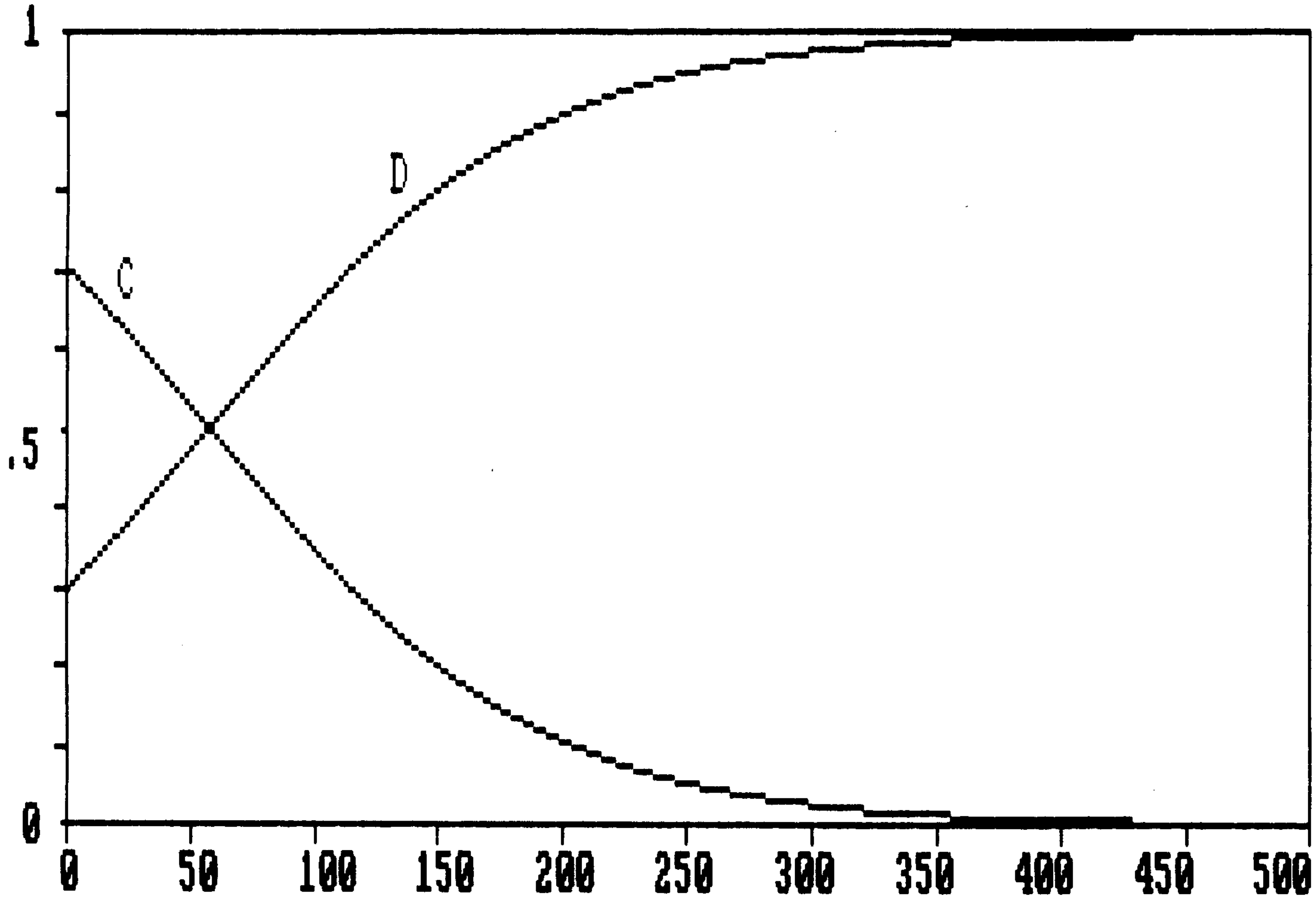


FIG. 1a

Prisoners' Dilemma: Evolution of strategies over 500 generations

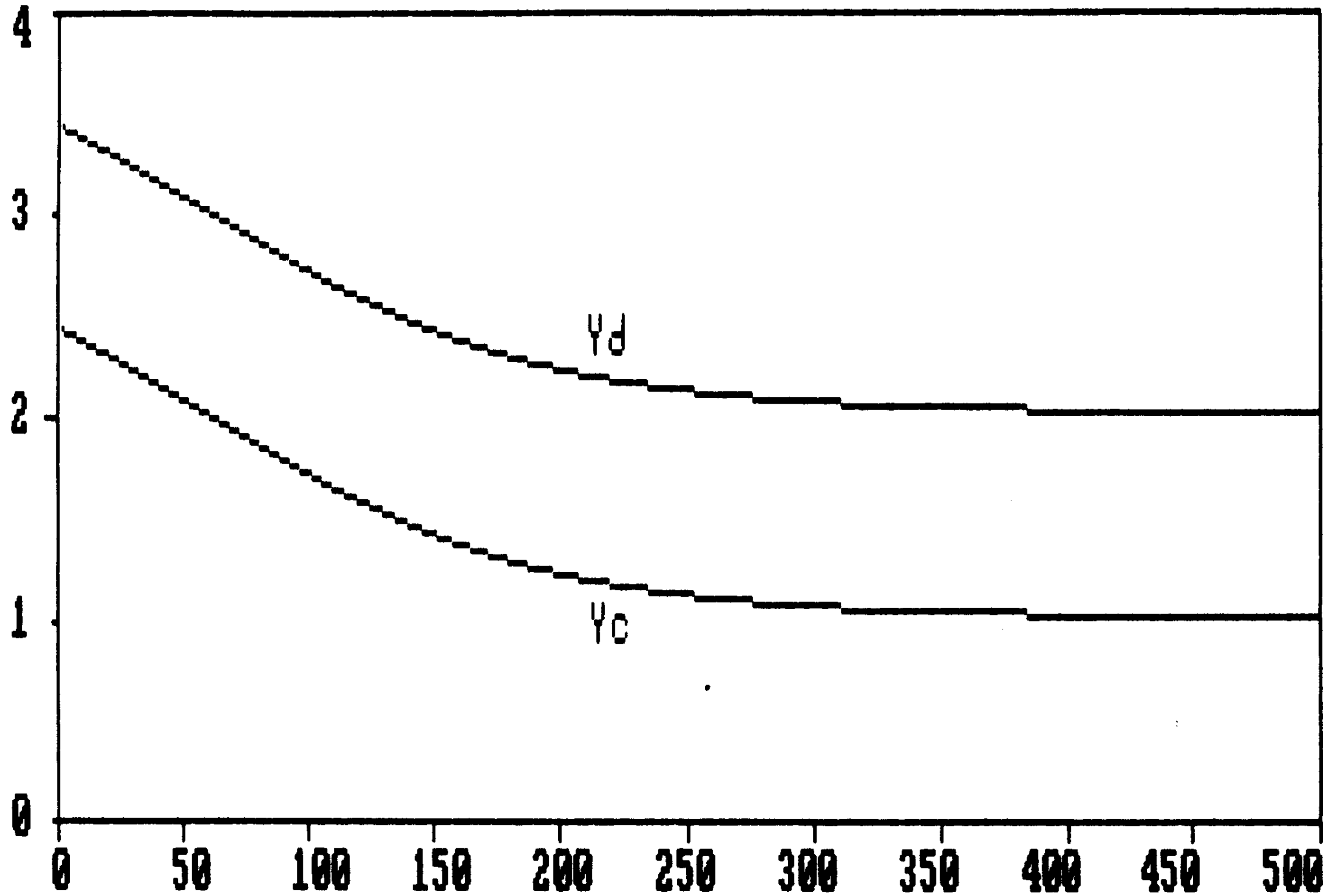


FIG. 1b

Prisoners' Dilemma: Evolution of yields over 500 generations

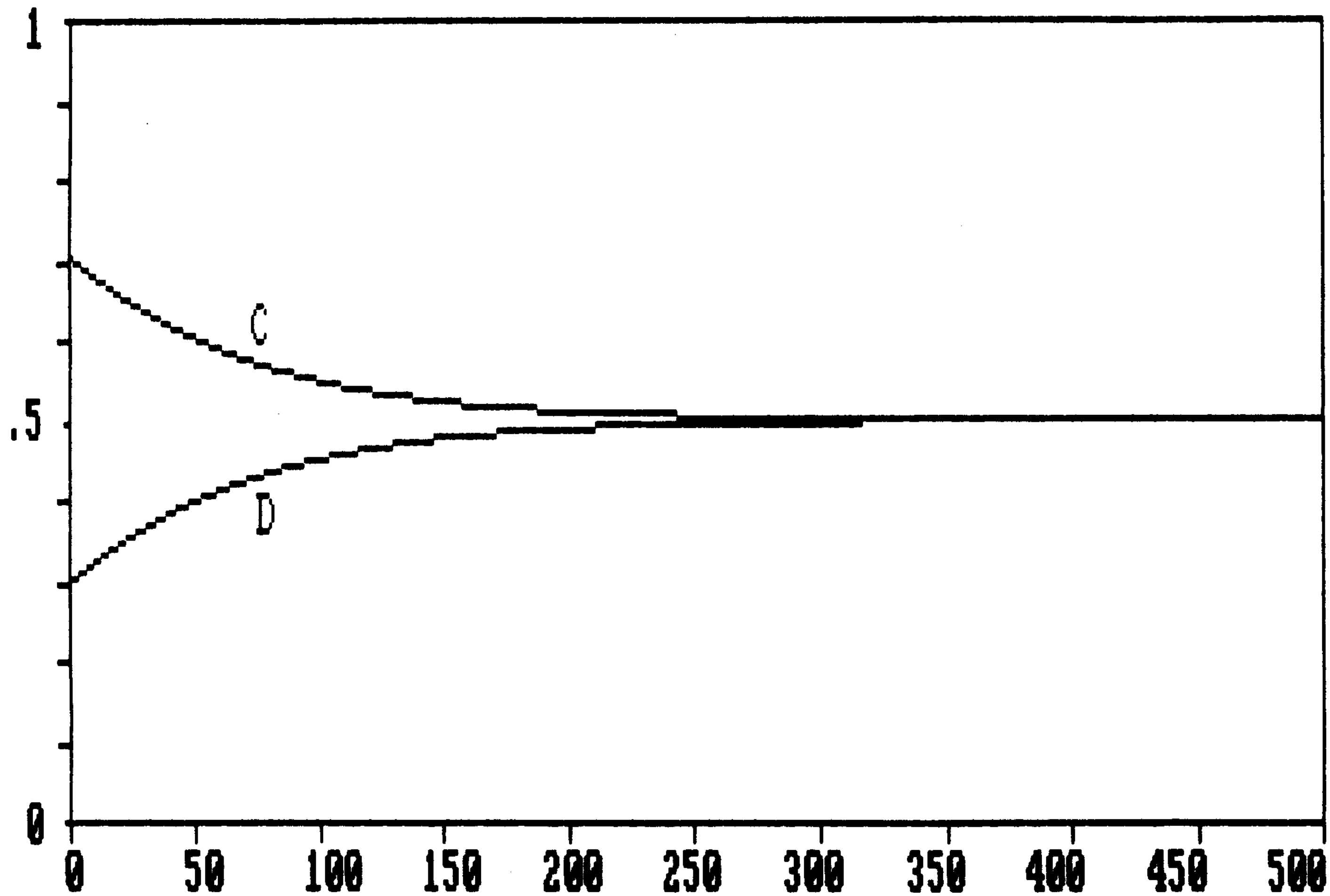


FIG. 2a

Chicken: Evolution of strategies over 500 generations

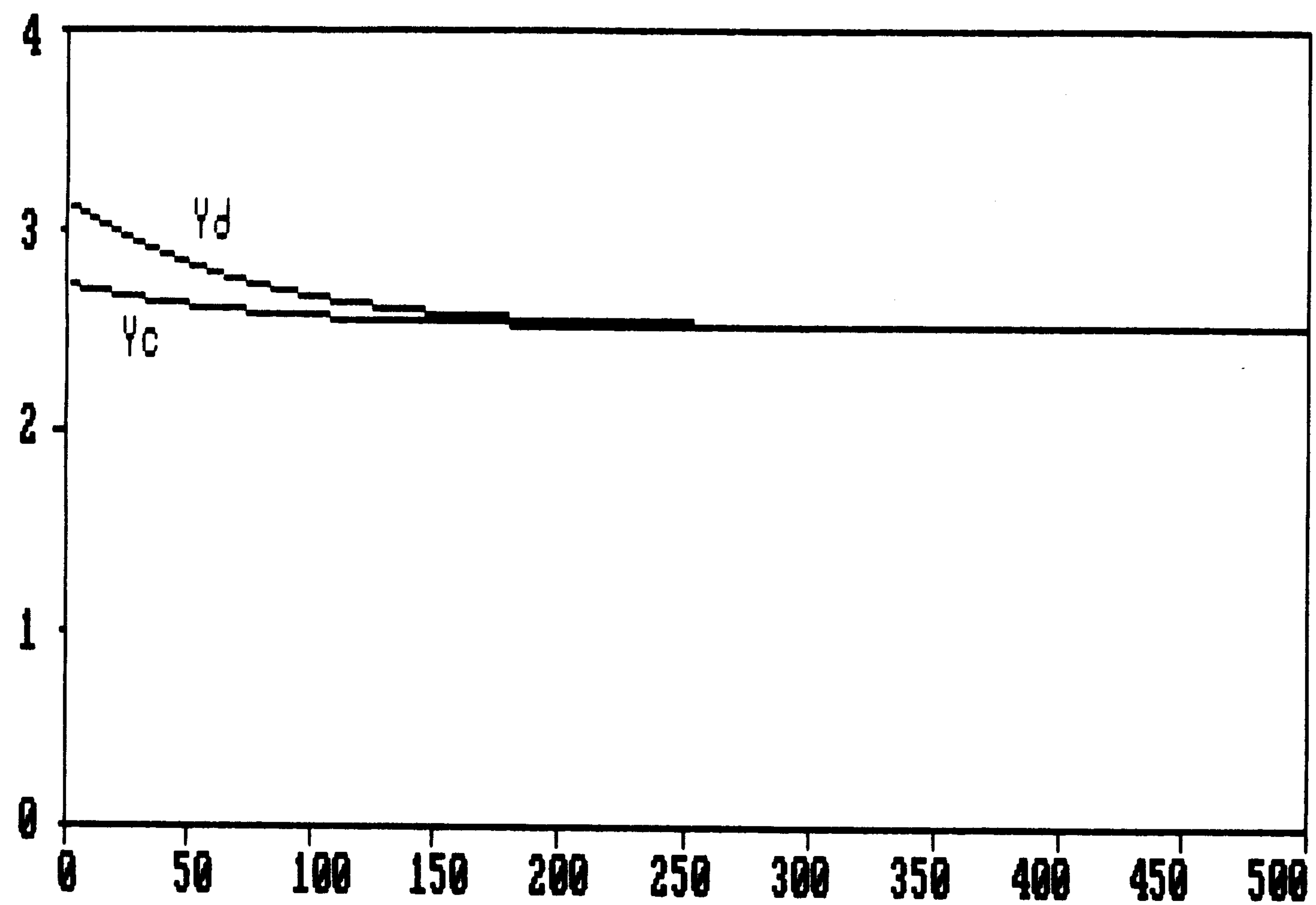


FIG. 2b

Chicken: Evolution of yields over 500 generations

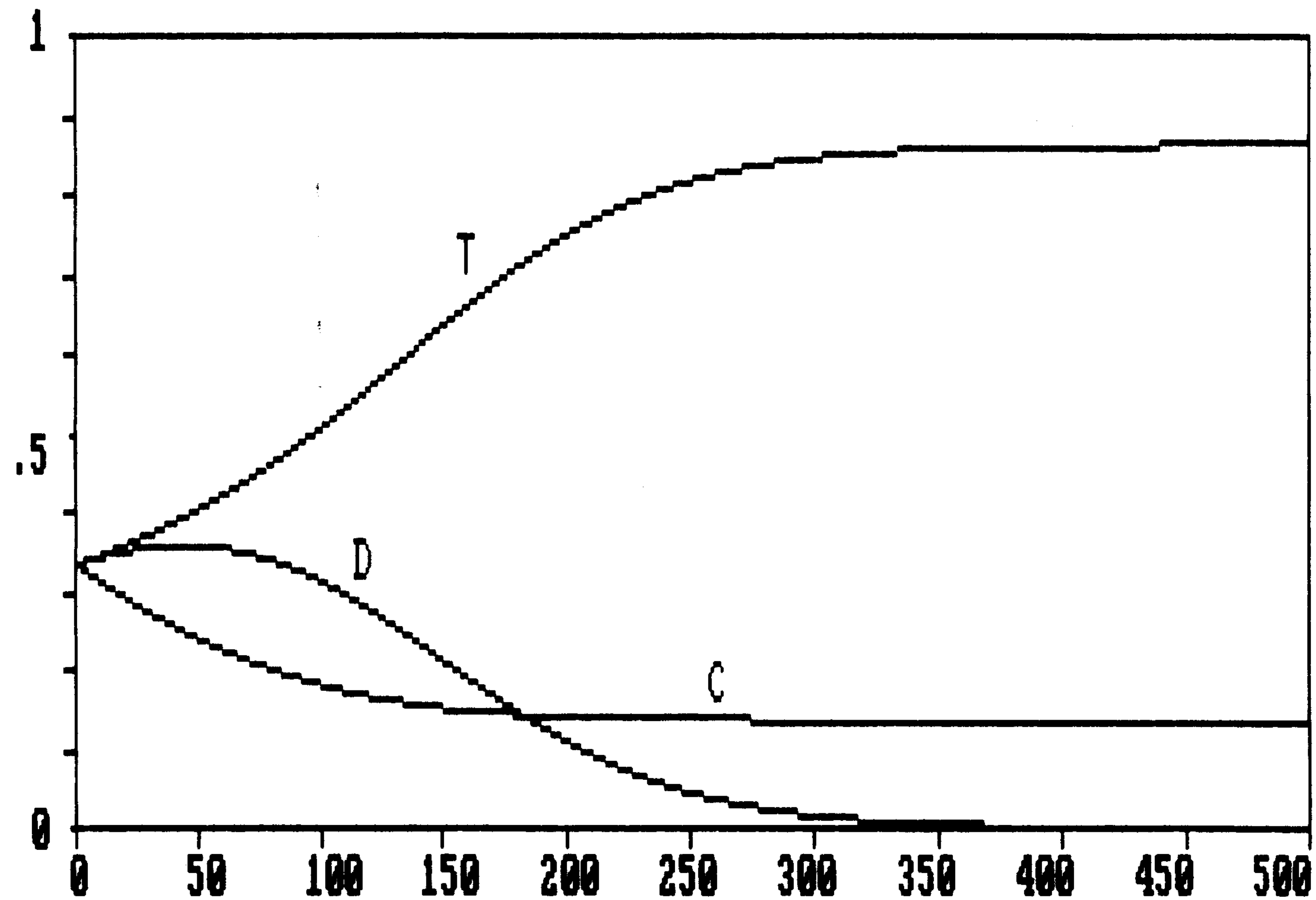
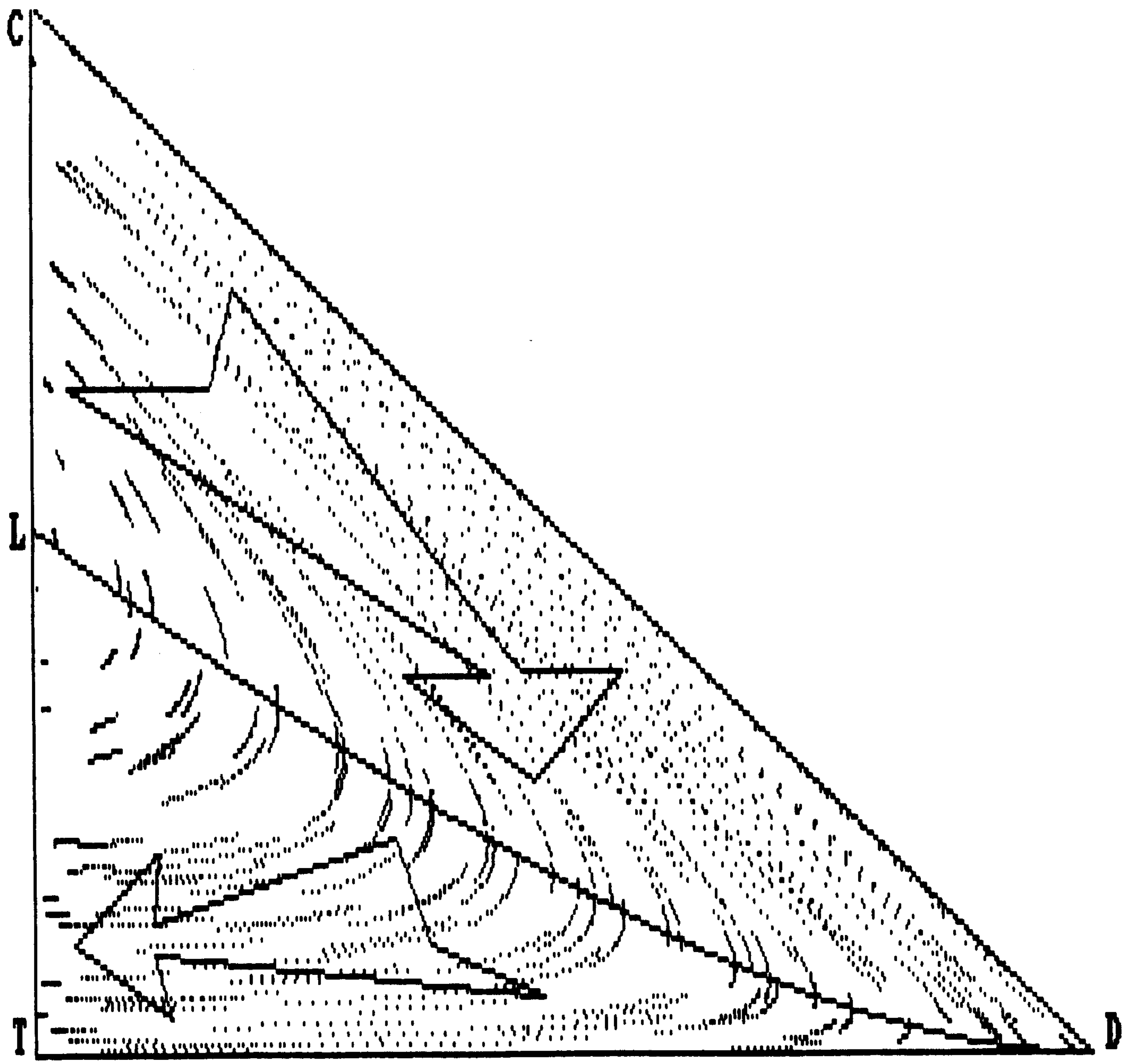
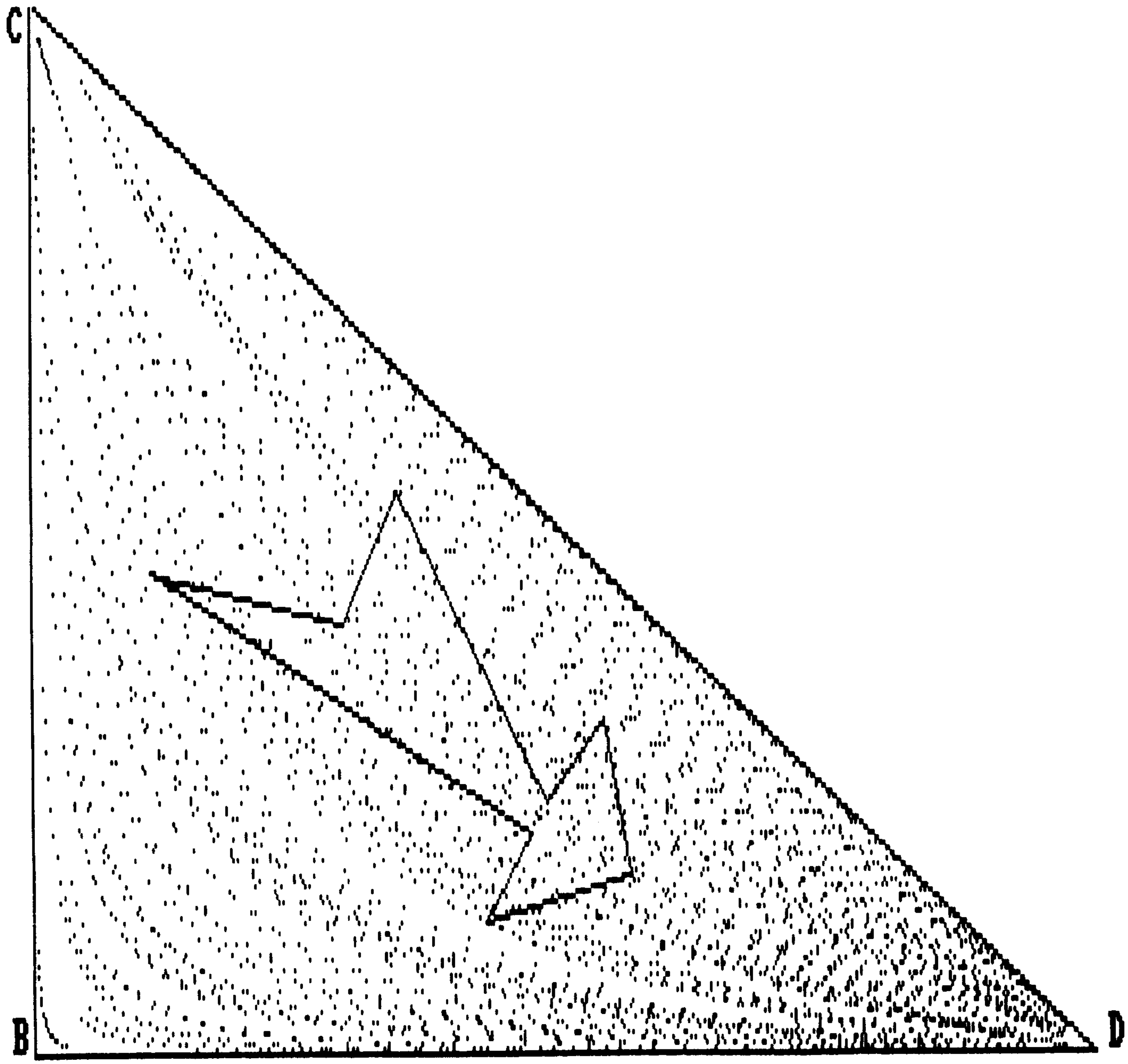


FIG. 3

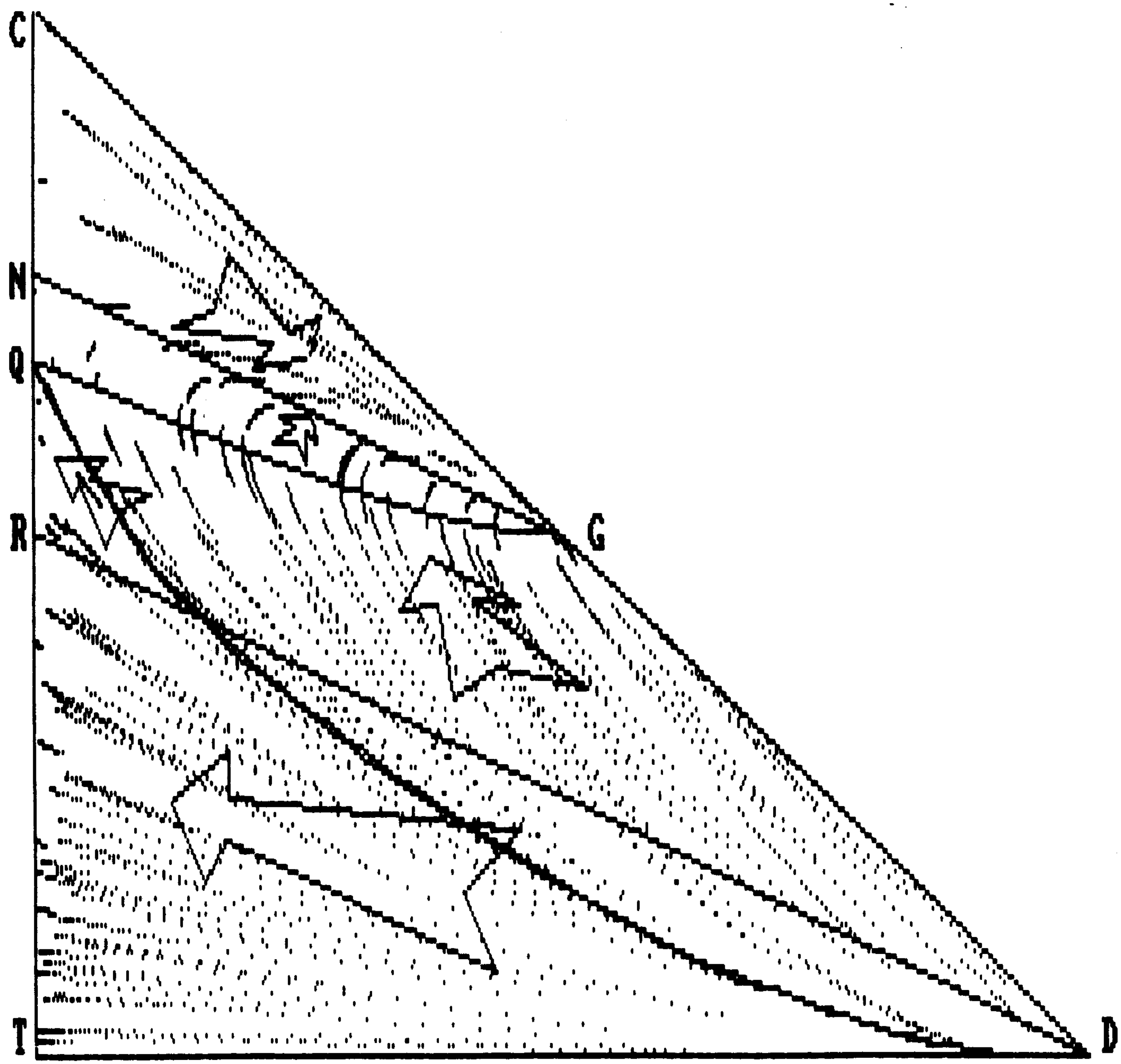
Prisoners' Dilemma plus TIT FOR TAT: Historical diagram



<u>Figure 4</u>: Prisoners' Dilemma plus TIT FOR TAT: Phase diagram



<u>Figure 5</u>: Prisoners' Dilemma plus BULLY: Phase diagram



<u>Figure 6</u>: Chicken plus TIT FOR TAT: Phase diagram

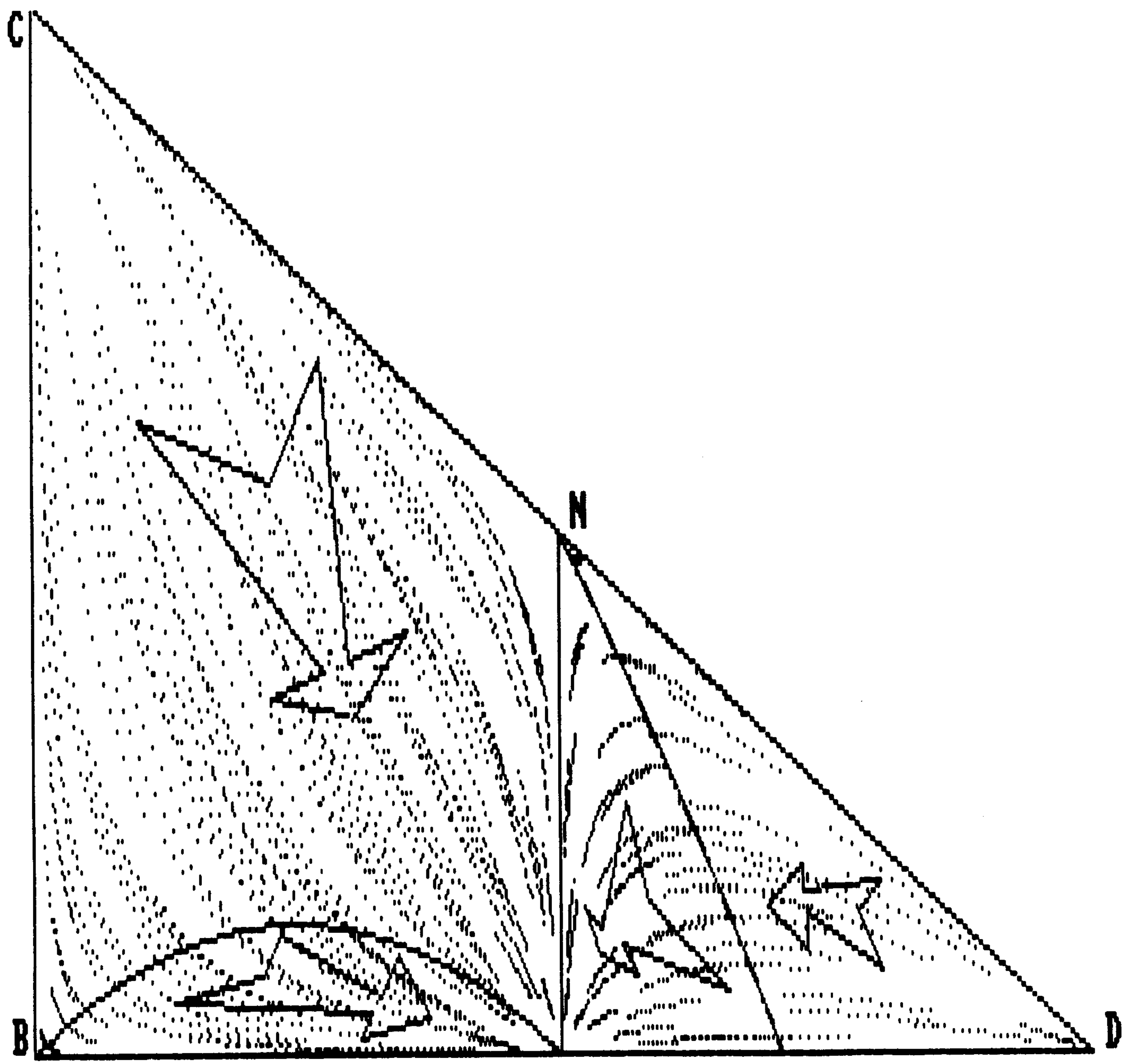
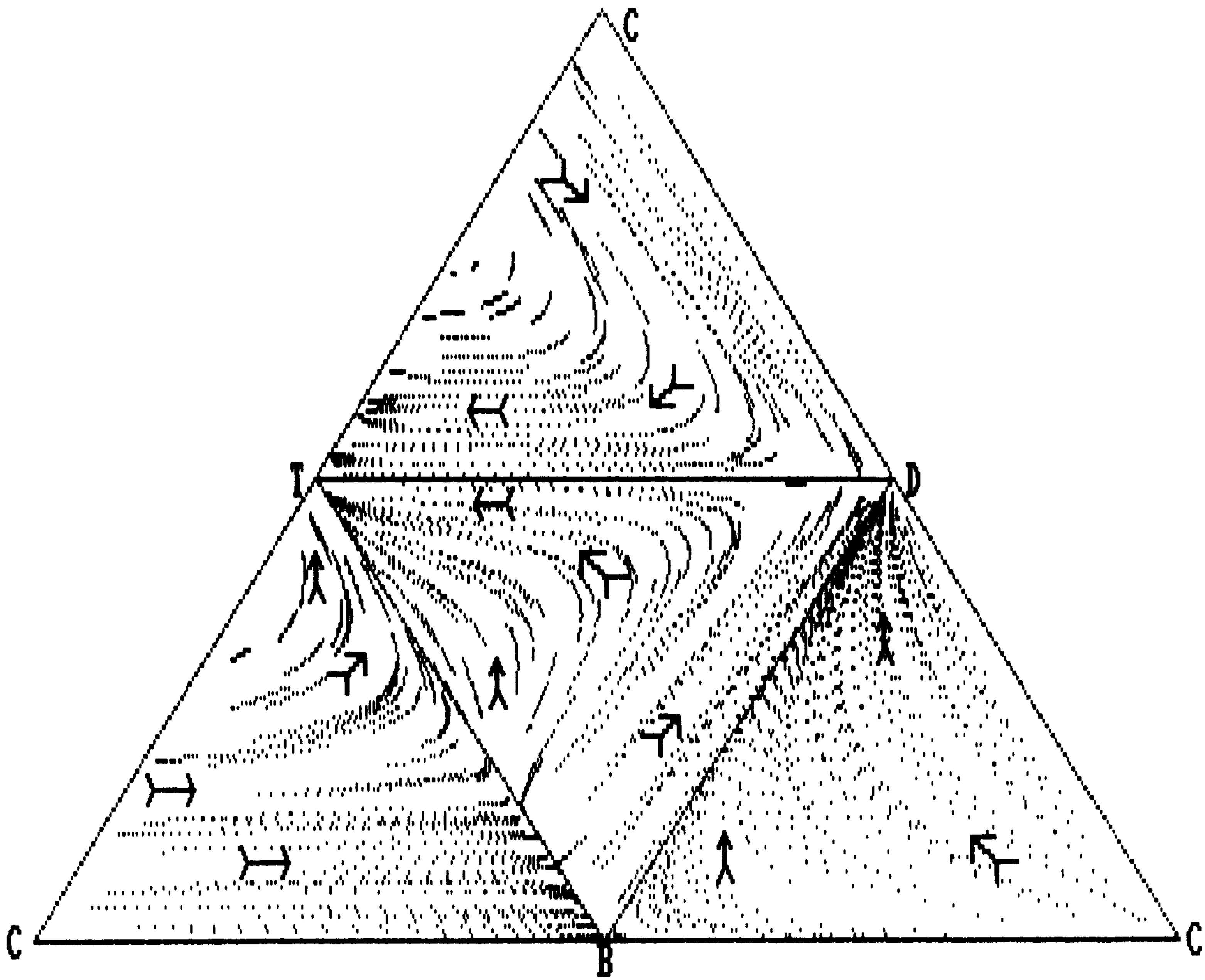
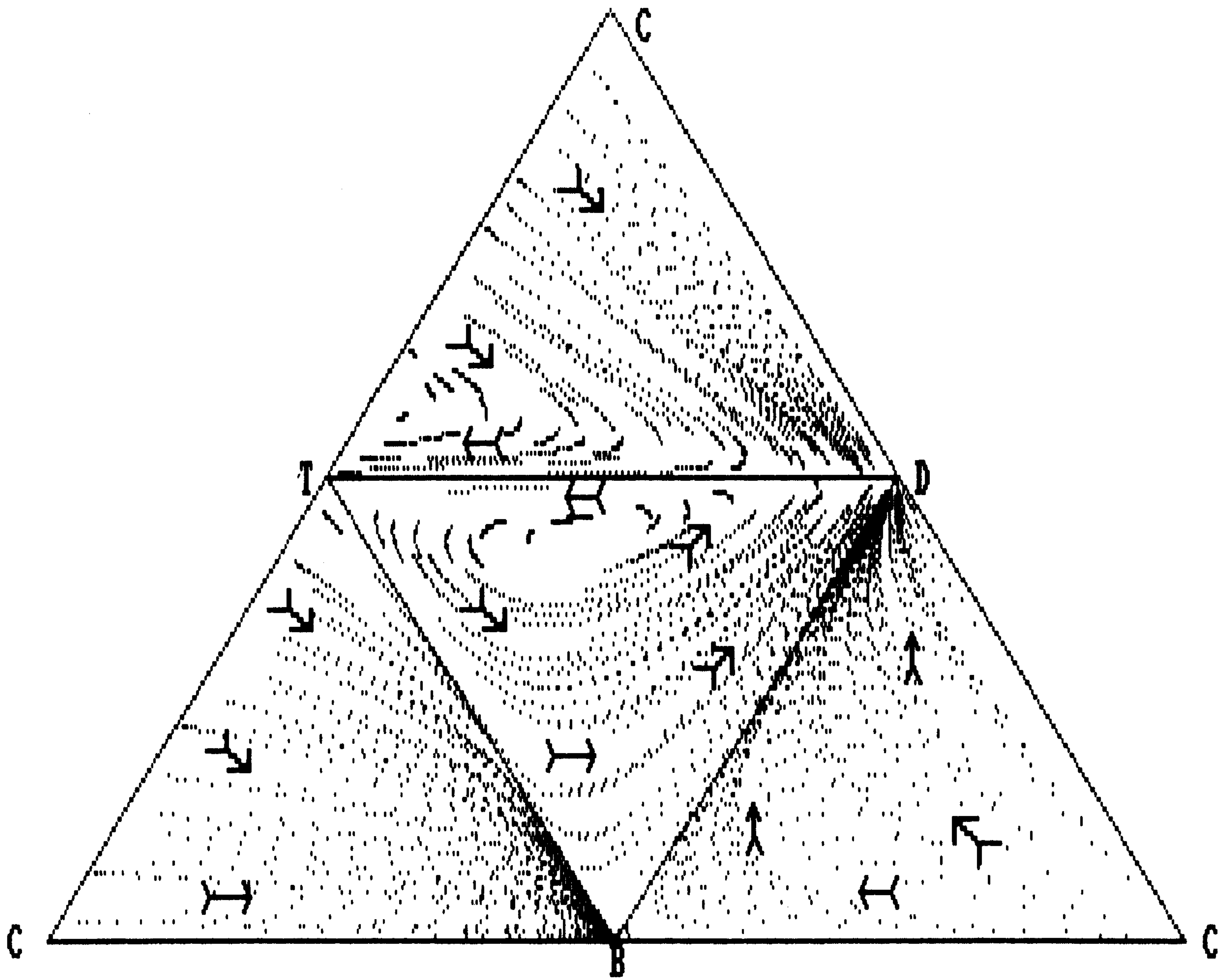


Figure 7: Chicken plus BULLY: Phase diagram



<u>Figure 8</u>: 4x4 Prisoners' Dilemma -- Symmetry-favoring payoffs



<u>Figure 9</u>: 4x4 Prisoners' Dilemma -- Asymmetry-favoring payoffs

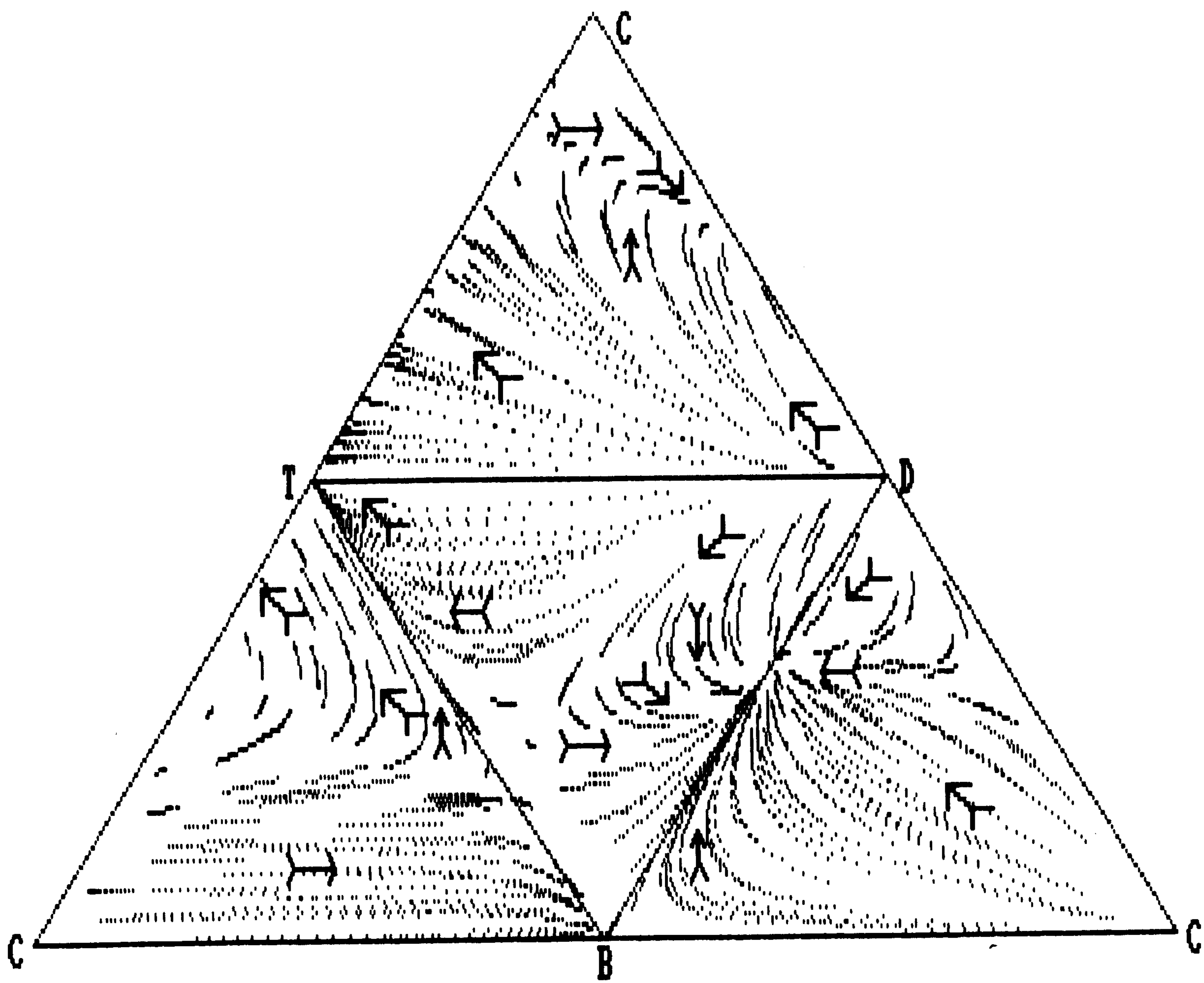
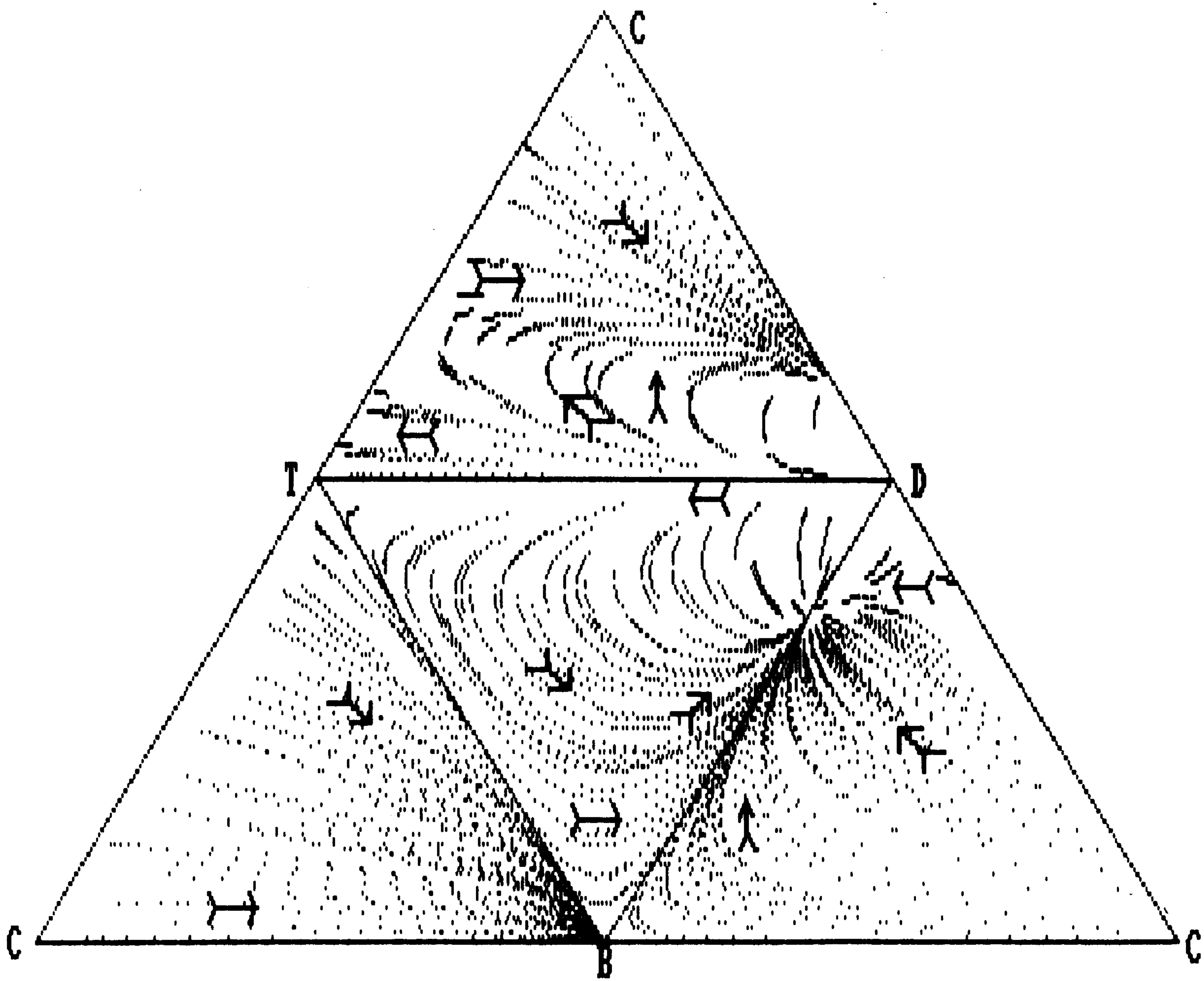


Figure 10: 4x4 Chicken -- Symmetry-favoring payoffs



<u>Figure 11</u>: 4x4 Chicken -- Asymmetry-favoring payoffs